University of Dundee

DOCTOR OF PHILOSOPHY

Personalised Medicine for Non-Alcoholic Fatty Liver Disease

Taylor, Alasdair

Award date:
2021

Personalised Medicine for Non-Alcoholic Fatty Liver Disease

Alasdair Taylor

For the degree of Doctor of Philosophy

School of Medicine

University of Dundee

Supervised by Professor Colin NA Palmer and Professor John F Dillon

February 2021

Table of Contents

List of Tables

List of Figures

Declarations

Declaration of candidate

I declare that this thesis is based on results obtained from investigations which I have carried out in the Division of Population Health and Genomics, University of Dundee, between December 2017 and November 2020 using funding provided by AstraZeneca. I declare that the entire thesis is my own composition. Any work other than my own is clearly stated in the test and acknowledged with reference to the relevant investigators or contributors. This thesis has never been presented previously, in whole or in part, for the award of any higher degree. I have consulted all the references cited within the text of this thesis.

Signed: _____[redacted]_____

Alasdair Taylor

Date: 2nd November 2020

Declaration of the supervisor

I certify that Alasdair Taylor has completed the equivalent of 9 terms of experimental research and that he has fulfilled the conditions of the University of Dundee, so that he is qualified to submit this thesis in application for the degree of Doctor of Philosophy.

Signed: Colin NA Palmer

Professor Colin N A Palmer

Date: 2nd November 2020

Acknowledgements

Many thanks to my supervisors Colin and John, who I've enjoyed working with immensely. Thank you to Colin for his excellent mentorship and ambition to do great science. Thank you to John for his expert guidance on all things hepatological and support throughout my PhD.

Thank you to everyone in the office; Moneeza, Cyrielle, Adem, Samira, Sundar, Abi, Yvonne, Enrique, Fred, Isobel, Ryan, Diane, David, Théo, Holly, Hollian, Laura, Ewan, Alex and Caroline.

Thank you to AstraZeneca for funding this project and to Carolina, Phil, Ben, Javier, Dirk, Slavé, and Margi for being excellent collaborators. Thank you to Stratified Medicine Scotland, and Diane and Lynn for their support. Thanks to everyone at the University of Dundee.

I would like to acknowledge the contributions of the following people in particular:

Thank you to my fiancée and best pal, Georgia.

Thank you to my mum and dad for always supporting me. Mum for her encouragement and Dad for all of the academic help throughout my education.

My brother and sister for their support.

Moneeza, whose wisdom and friendship has been crucial to this whole operation.

Cyrielle, for all the GWAS help and being my AZ buddy.

Adem, for the academic guidance and being a friend.

Caroline and Isobel for all of their help and excellent chat.

My friends Shaina, Ben, Lauren, Jack, Luke, Katie, Chris, Josh, Kirsty, Dave, Sorcha, Jamie, Kathryn, Zoe, and Eilidh.

Dedication

For Georgia.

For my mum, dad, brother and sister.

Abstract

Non-Alcoholic Fatty Liver Disease (NAFLD) is a chronic and dangerous condition which has grown in prevalence over recent decades due to increasing rates of obesity, affecting roughly a quarter of adults globally. This thesis develops a useable NAFLD definition which can be applied to large retrospective data sets consisting of medical records for selected cohorts. The morbidity and mortality associated with NAFLD is investigated, especially related to extrahepatic cancer. This thesis also aims to identify genetic modifiers of NAFLD risk in Scottish and South Indian populations.

Data from three retrospective Scottish cohorts with electronic health records (EHRs) were analysed in the current thesis. These were the GoDARTS, SHARE and Tayside and Fife Diabetics cohorts. Genotypic data was available for a number of patients in GoDARTS and SHARE. Data from the Dr Mohan's Diabetes Speciality Clinic (DMDSC) were also analysed, which consisted of clinical measurements from clinic visits, and genotypic data.

An accurate and practical NAFLD definition based on two raised ALT levels was developed, which had a sensitivity of 97.4% in the GoDARTS cohort. This definition was used for subsequent NAFLD analyses. Patients with NAFLD experienced significantly more hospital admissions, and earlier death than those without NAFLD. We found NAFLD is associated with increased risk of cancer incidence, and cancer death, which accounts for a large portion of the excess morbidity and mortality seen in NAFLD patients. GWAS analyses revealed that *PNPLA3* rs738409 is a major genetic risk factor in both Scottish and South Indian populations. Candidate gene studies revealed variants associated with endothelin function and with GLP1 receptors had significant effects on NAFLD.

This thesis applies an accurate and novel NAFLD definition to retrospective cohorts with EHRs, and found increased morbidity and mortality in individuals with this phenotype. A

large proportion of this is explained by increased cancer incidence and cancer death seen in these individuals. A number of genetic risk factors for NAFLD are described, including novel loci in endothelin and GLP1R related genes.

1 Introduction

1.1 Introduction to Non-Alcoholic Fatty Liver Disease

Non-Alcoholic Fatty Liver Disease (NAFLD) is a chronic condition which affects roughly 25.2% of the adult population globally, and is now the leading cause of liver disease.[1] Prevalence of NAFLD has risen in recent decades, and continues to rise due to increased rates of obesity.[2] There is currently no recommended pharmacological intervention for NAFLD.[3] The high prevalence of this disease and associated morbidity and mortality make NAFLD a key research priority, with the aims of improving assessment of NAFLD risk and informing drug discovery.[4]

1.2 NAFLD Overview

NAFLD is a multi-stage disease which can have deadly consequences.[5] These stages are shown in figure 1-1 below.[6] Though each of these stages is associated with NAFLD, they do not necessarily occur in a linear fashion, as some patients may have fibrosis with no steatohepatitis for example.

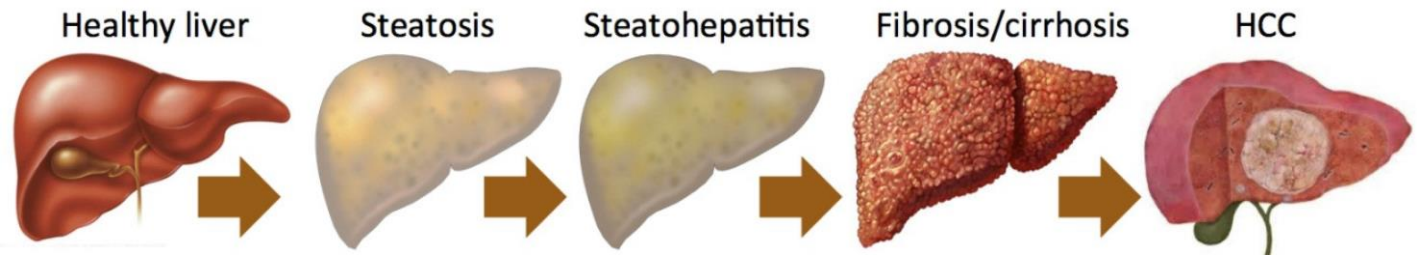

Figure 1-1 Stages of NAFLD Progression

The first stage of NAFLD is simple steatosis, where macrovesicular steatosis is present in >5% of hepatocytes.[7] Steatosis is the fatty infiltration of hepatocytes, without inflammation or fibrosis present. Triglyceride droplets form in the cytoplasm of the liver cells in large quantities, outside of normal ranges. There are many mechanisms involved in the build-up of steatosis, but these mainly involve dysregulation of lipid metabolism and increased availability of lipids, due to conditions such as obesity and diabetes.[8] Simple steatosis is the

most common form of NAFLD, with a minority of patients progressing to more severe stages per year of disease.[7]

Non-Alcoholic Steatohepatitis(NASH) occurs where NAFLD causes inflammation in the liver cells.[9] It is estimated that 25% of patients progress from simple steatosis to NASH over a three year period.[5] NASH is characterised by inflammation of the hepatocytes due to fatty degeneration, though this inflammation does not correlate closely with the severity of steatosis. This causes injury to the hepatocytes and leads to cell death. The triglycerides which form the steatosis are not toxic to hepatocytes, therefore it is thought that lipid intermediates and oxidative stress are likely to be the cause of the toxicity and therefore damage to the cells.[10,11] NASH is reversible, and improvements in NASH have been reported in patients following bariatric surgery.[12] NASH is most accurately diagnosed by biopsy, and has distinct properties compared with NAFLD. Histologic features of NASH are mainly inflammation, hepatocyte ballooning, and Mallory-Denk bodies.[13] There are however issues with sampling variability with biopsies for NASH diagnosis, as lesions can be spread unevenly throughout the liver, and inter-observer variability can affect diagnosis.[14]

Hepatic fibrosis can occur in those with NAFLD.[5] This is characterised by the deposition of extracellular matrix in the parenchyma of the liver.[15] Fibrogenesis in the liver is usually the result of a chronic wound healing process, in response to damage caused by NASH in the case of NAFLD.[15] As part of the response to repair the liver, fibrotic tissue is produced by hepatic stellate cells. (HSCs) These cells are usually in a dormant state, and have a role in the storage of vitamin A.[16] When the wound healing processes are activated in the liver however, HSCs are activated and proliferate. They produce large amounts of extracellular matrix, which cause major structural changes in the liver. It is estimated that 20% - 30% of patients with NAFLD progress to fibrosis over three years.[5] Patients with fibrosis progress at one

fibrosis stage per decade.[17] Initially thought to be irreversible, there is now some evidence to show that fibrosis can be reversed to some extent.[18]

Cirrhosis occurs in extreme cases of fibrosis, where a portion of the liver is replaced by the scar tissue formed by fibrosis, and regenerative nodules of hepatocytes form.[19] This drastically changes the structure of the liver, causing reductions in function, and portal hypertension.[20] In cirrhosis, regenerative nodules of hepatocytes form between strands of scar tissue (septa), as the liver tries to repair itself.[21] Cirrhosis is extremely dangerous, and can lead to liver failure. It is thought that up to 38% of patients with fibrosis progress to cirrhosis.

Hepatocellular carcinoma (HCC) is a complication which can occur in cirrhotic livers. It is the 6th most common cancer and is a major cause of cancer death globally.[22,23] A number of pathological changes occur in cirrhotic livers to provide an environment conducive to neoplasia.[24] Inflammation is thought to play a large role in this, influencing several genetic and epigenetic transformations leading to neoplasia.[24] Regenerative nodules in the liver often show dysplasia, leading to HCC. A small but significant portion of patients with NAFLD develop HCC. The proportion is estimated to be 2.4% - 12.8% in patients with NASH over a 3.2–7.2 year period.[5] Around 20% of HCC occurs in non-cirrhotic patients, though a certain amount of this may be due to sampling variability in biopsies and missed cirrhosis diagnoses.[23]

Though NAFLD is common, it is underdiagnosed in clinical settings.[25] This is due to the challenges of diagnosing NAFLD, which is often asymptomatic.[26] There is evidence that NAFLD is a key contributor to cryptogenic cirrhosis.[27] The full scale of the risk posed by NAFLD, as well as its epidemiology cannot be understood without accurate diagnostic techniques. The gold standard technique is biopsy, though this invasive and can have complications which can lead to death in some cases.[28,29] Imaging and biochemical methods

of diagnosis are more commonly used, though vary in accuracy and the validity of some measures is debated.[30,31]

The concept of NAFLD dates back to the early 80's when Ludwig et al. first described NASH in a landmark paper; though NAFLD research was not undertaken intensively until later in the 1990's.[32] Prior to this, it was assumed that all patients who reported zero or low alcohol use, and had fatty livers were dishonest about their alcohol intake. NAFLD and its alcoholic related counterpart, alcoholic liver disease (ALD) have a large amount of pathophysiological overlap, and are indistinguishable in many ways.[33] NAFLD is defined as the presence of fatty infiltration of the liver in the absence of excess alcohol intake or other cause of liver disease. Therefore, upper limits of alcohol intake are set to prevent ALD being categorised as NAFLD. Other causes of liver disease including virological and immunological insults must also be ruled out too to diagnose NAFLD.

The use of "Non-Alcoholic" in NAFLD and NASH was used as the definition was being presented as an alternative to ALD, and compared to it.[32] In recent times, the limitations of this definition have been discussed an updated terminology suggested. The term metabolic associated fatty liver disease or "MAFLD" has been proposed to more accurately describe the condition, as it provides a description of the aetiology of the disease.[34] The author thought this to be important as the condition can coexist with a number of other liver diseases, including ALD, which is an important factor in understanding the disease. Though "MAFLD" is a more descriptive and appropriate name, to maintain consistency and clarity, the term NAFLD will be used throughout this thesis as this is the terminology used in almost all previous literature and is still widely used.

1.3 NAFLD Pathogenesis

A number of factors influence the pathogenesis of NAFLD, shown below in figure 1-2 from the 2016 paper by Buzzetti et al, "The multiple-hit pathogenesis of non-alcoholic fatty liver disease (NAFLD)".[35]

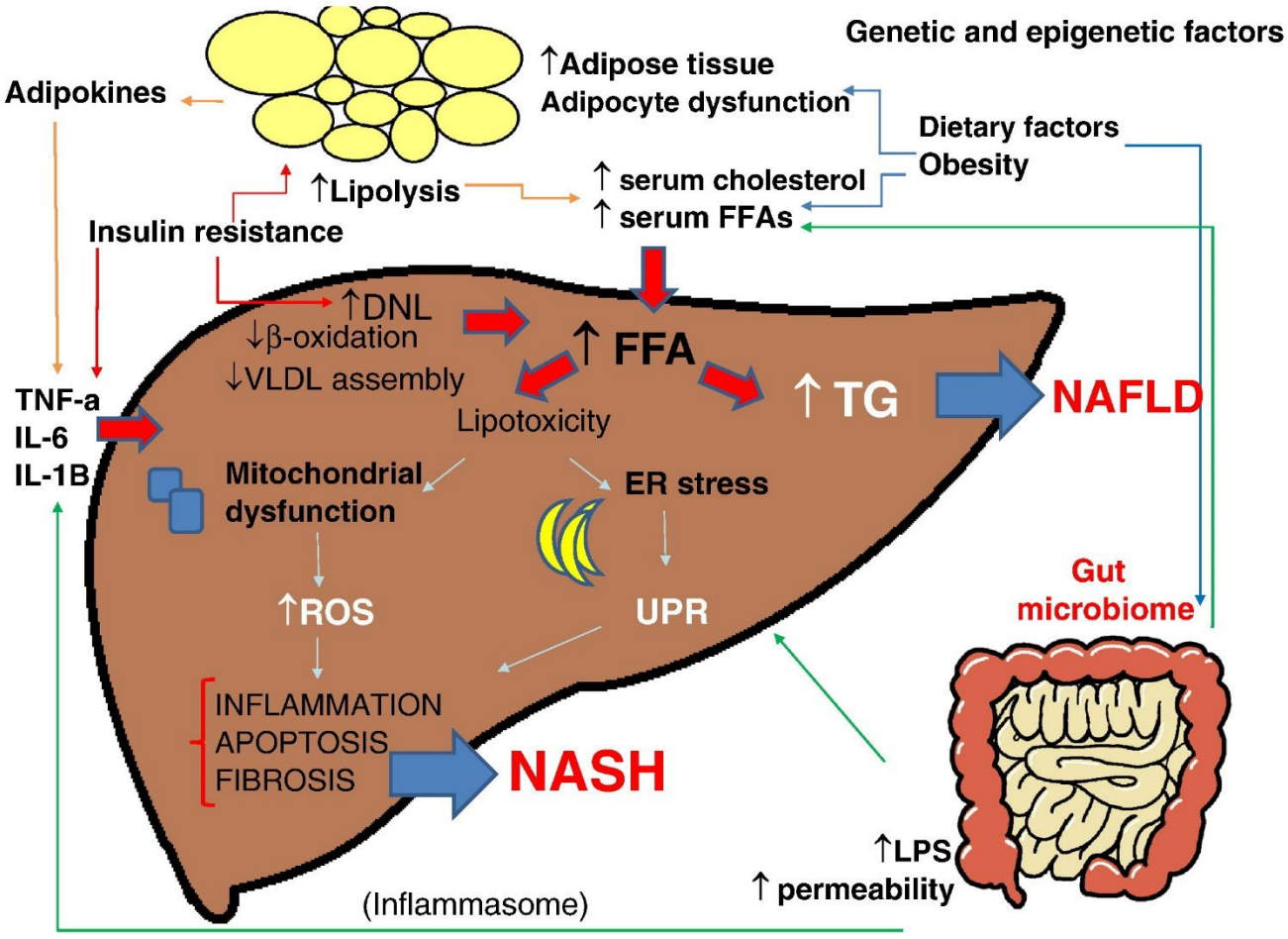

Figure 1-2 NAFLD Pathogenesis

This diagram does not show all pathways involved in the development of NAFLD, NASH and fibrosis, but includes many key features. Firstly dietary factors such as high carbohydrate and fat intake increase risk of obesity.[36] Outside of the liver this leads to increased adipose tissue and dysfunction. The adipose tissue releases inflammatory adipokines which can lead to mitochondrial dysfunction in the liver, and inflammation.[37] In the gut, the microbiome is affected, and the gut becomes more porous to factors including endotoxins.[38] Obesity and insulin resistance increase the free fatty acid (FFA) and cholesterol content of the blood, and

much of these are absorbed by the liver. Hepatic insulin resistance increases de novo lipogenesis, which further increases the total FFA content of the liver. Much of this excess FFA is converted to triglycerides, stored in the hepatocytes.[35] This excess triglyceride storage is steatosis. Triglycerides themselves are not hepatotoxic, however it is thought that when the liver is no longer able to store FFA's as triglycerides, the FFA's can be harmful to hepatocytes. FFA's which are not converted to triglycerides can increase risk of mitochondrial dysfunction and endoplasmic reticulum stress. These processes increases number of reactive oxygen species and unfolded protein response respectively.[39,40] These processes lead to inflammation which characterises NASH. Inflammatory adipokines also contribute to inflammation of the cell. This cellular damage leads to cell death, and subsequent injury response in the form of fibrosis, from activated hepatic stellate cells.[13] This fibrosis can take over large portions of the liver, becoming cirrhosis.

A number of other factors contribute to the pathogenesis of NAFLD, many influenced by genetics. There is significant heterogeneity in the clinical presentation of patients with NAFLD and NASH, as some can have normal BMI for example.[41] In these individuals, it is likely that dysfunction has occurred in a process downstream of obesity.

1.4 NAFLD in Clinical Practice

NAFLD is in the first instance suspected by a physician based on overall clinical presentation based on risk factors including age >50, being male, obese, high cholesterol and other lipids. NAFLD is usually diagnosed first via elevated liver function tests.[42] Alanine transaminase (ALT) is an important marker used in the preliminary diagnosis of NAFLD, and is covered in more detail in the following chapter. NAFLD can also be suspected in the first instance by incidental findings in ultrasonography.[42] The majority of patients who are diagnosed with NAFLD are obese, and often have other comorbidities such as T2DM and hypertension. Though progression to more severe stages in NAFLD is relatively uncommon, the prevalence

of the disease and association with other conditions such as T2DM make diagnosis of NAFLD important. No pharmacological treatments are currently suggested, though lifestyle modification through diet and exercise have been found to lower liver fat content. [43] NICE guidelines state that lifestyle modification is the only evidence based method for treatment of NAFLD.[44] Physicians encourage these lifestyle modifications in patients, and target the lowering of liver enzymes and weight loss.[42] This management policy for NAFLD patients is sufficient for most, though severe cases can be referred to hepatologists for closer monitoring.

Significant debate exists as to whether blanket screening for NAFLD should be carried out in primary care.[45] The key concern is balancing practicality, cost and invasiveness against the clinical utility and improvement to patient care gained by screening. Screening for NAFLD in the population is not recommended in the NICE guidelines, as there is insufficient evidence to suggest this would be worthwhile.[44] Nascimbeni et al. reviewed NAFLD screening guidance from a number of international and national hepatological societies and found none recommended screening for NAFLD in the general population .[46] Most guidance recommends targeted screening for those considered to be at greater risk of NAFLD. [45,47] Patients with features of metabolic syndrome, and insulin resistance in particular are targeted, and some studies have found these techniques to be cost effective.[48]

Others suggest that screening for NAFLD is unjustified, and wastes resources and time. The majority of NAFLD does not progress to fibrosis or HCC, and appears to be benign in many, therefore treatment of NAFLD many not provide benefit.[49] Lifestyle modification is at present the main intervention, and given the fact obesity, T2DM and other features of metabolic syndrome are so highly correlated with NAFLD it is likely patients will be advised to lose weight by their clinician regardless of NAFLD status. Rowe suggests in his paper "Too much medicine: overdiagnosis and overtreatment of non-alcoholic fatty liver disease",

that screening for NAFLD even in at-risk groups is problematic as it risks incursion of harm to patients who are falsely diagnosed.[50] Other such as Malnick also suggest over treatment of NAFLD is harmful, and that the considerable amount of money being spent of pharmacological interventions would be better placed improving public health facilities and helping enable lifestyle modifications.[51] Others have presented similar economic arguments, suggesting it is not cost effective to screen for NAFLD. It may be cheaper and more effective to advocate weight loss and lifestyle modification than to screen and treat NAFLD. The risk with this approach however is that lean individuals and those without the traditional comorbidities of NAFLD may not be diagnosed and therefore progress to more serious stages of NAFLD without intervention.

NAFLD is common, but it is underdiagnosed in clinical settings.[25] This is due to the challenges of diagnosing NAFLD, which is often asymptomatic.[26] Aside from improving NAFLD screening and diagnosis techniques to improve patient care, this will also aid scientific research, the full scale of the risk posed by NAFLD, as well as its epidemiology cannot be understood without accurate diagnostic techniques. This will allow GWAS and other genetic analyses to be conducted to find further genetic risk factors for NAFLD. One of the gold standard techniques is biopsy, though this invasive and can have complications which can lead to death in some cases.[28,29] Imaging and biochemical methods of diagnosis are more commonly used, though vary in accuracy and the validity of some measures is debated.[30,31] Different methods of screening may also be employed to increase practicality and reduces costs, including applying NAFLD diagnosis algorithms to retrospective medical records to determine those at risk of NAFLD. This method is used in the current thesis.

The concept of NAFLD dates back to the early 80's when Ludwig et al. first described NASH in a landmark paper; though NAFLD research was not undertaken intensively until later in the 1990's.[32] Prior to this, it was assumed that all patients who reported zero or low

alcohol use, and had fatty livers were dishonest about their alcohol intake. NAFLD and its alcoholic related counterpart, alcoholic liver disease (ALD) have a large amount of pathophysiological overlap, and are indistinguishable in many ways.[33] NAFLD is defined as the presence of fatty infiltration of the liver in the absence of excess alcohol intake or other cause of liver disease. Therefore, upper limits of alcohol intake are set to prevent ALD being categorised as NAFLD. Other causes of liver disease including virological and immunological insults must also be ruled out too to diagnose NAFLD.

The use of "Non-Alcoholic" in NAFLD and NASH was used as the definition was being presented as an alternative to ALD, and compared to it.[32] In recent times, the limitations of this definition have been discussed an updated terminology suggested. The term metabolic associated fatty liver disease or "MAFLD" has been proposed to more accurately describe the condition, as it provides a description of the aetiology of the disease.[34] The author thought this to be important as the condition can coexist with a number of other liver diseases, including ALD, which is an important factor in understanding the disease. Though "MAFLD" is a more descriptive and appropriate name, to maintain consistency and clarity, the term NAFLD will be used throughout this thesis as this is the terminology used in almost all previous literature and is still widely used.

1.4.1 Prevalence of NAFLD and Clinically Significant NAFLD

There are many estimates of the prevalence of NAFLD globally. [25,52–54] Younossi et al. meta-analysed 86 studies from 22 countries, and calculated a global average of 25.24% prevalence of NAFLD in adults. Significant heterogeneity was seen between countries and ethnicities, though it is unclear how much of this is due to differences in case ascertainment and reporting.[55] Particularly high prevalence has been found in South American countries; as high as 30.45% based on ultrasound findings.[49] Lower prevalence overall is seen in China, estimated to be around 15%, although this varies by region and is as high as 27% in some

urban populations. These rates are highly correlated with lifestyle factors which can vary drastically between different communities.

The overall prevalence of NAFLD is high, though the more advanced stages are rarer. Younossi et al. estimate global prevalence of NASH to be 3% in their meta-analysis.[1] They also calculated that approximately 41% of patients with NASH will progress to fibrosis. It is clear from these figures that a small but significant portion of individuals progress to the more harmful stages of NAFLD.

Simple steatosis in NAFLD is commonly described as a benign condition in the absence of fibrosis and cirrhosis.[56] In NAFLD, advanced fibrosis is the strongest predictor of disease specific mortality.[57] Some suggest the majority of NAFLD is not clinically significant, as it is unlikely to progress and become harmful, and is relatively harmless on its own. There is growing evidence that this is not the case however, as significant excess mortality has been found in those with even simple steatosis, and increased incidence of extrahepatic cancer is a major cause of this.[58–60] Allen et al. found those with NAFLD live on average 4 years shorter than those without, and cancer is a large part of this.[58] This is covered in detail in a later chapter of this thesis. Further to this increased mortality, NAFLD is also a risk factor for development of T2DM.[61]

These factors complicate the notion of clinically significant NAFLD, as non-progression to NASH and fibrosis does not necessarily mean that steatosis is not causing harm. The risk to health conferred by NAFLD must be assessed in the context of the intra and extrahepatic conditions it is associated with. It is therefore important that further study of NAFLD should aid prediction of these conditions, and determine the cause of these associations.

1.5 NAFLD Risk Factors

Despite association with increased age and body fatness, NAFLD can affect almost anyone, with paediatric and geriatric cases reported in significant numbers, and in every age group in between.[62] It can also affect lean individuals as well as those who are overweight and obese.[41] Despite this, there are a number of risk factors associated with NAFLD which increase risk of steatosis and the subsequent stages of NAFLD.[63]

The key risk factor for NAFLD is excess body fat.[37] Studies suggest that NAFLD is over four times more prevalent in obese patients.[64] In obese patients, the release of free fatty acids (FFAs) from visceral fat is increased which leads to an increase in uptake of FFAs by the liver leading to steatosis.[65] Though overall fatness predicts NAFLD, central obesity is more of a risk factor for NAFLD.[66] Visceral fat is metabolically active, and insulin resistance forms a large part of the relationship between NAFLD and obesity, as steatosis severity correlates with hepatic insulin resistance.[65] Reduction in bodyweight has been found to be effective in reducing NAFLD.[67] Exercise has been shown to reduce liver fat, independently of weight loss.[68] An association between low levels of physical activity and NAFLD has also been shown.[69] It is however possible for lean individuals (BMI < 25kg/m^2) to have NAFLD.[41] Estimates suggest that prevalence of lean NAFLD is as high as 7% in the USA, and 19% in parts of Asia.[70]

Age is a non-linear risk factor for NAFLD.[62] Age is associated with increased steatosis, as well as a number of other changes in the liver, including: increased fibrosis, decreased regeneration ability, increased inflammatory changes and increased oxidative stress. These factors increase the damage to the liver caused by NAFLD. Most studies suggest an inverse U shaped curve of NAFLD risk versus age, with peak risk around 50 to 60 then a decline in risk thereafter.[19,71,7272] Men are likely to get NAFLD 10 years earlier than women, but being post-

menopausal increases risk in women.[62] Paediatric NAFLD is relatively less common than NAFLD in adults, though prevalence is rising with increased paediatric obesity.[58]

Type 2 diabetes is a large risk factor for NAFLD.[73] In the general population, the rate of NAFLD globally is ~25%, but in those with T2DM the prevalence is much higher, with estimates at 68% in European cohorts.[55,74,75] NASH, fibrosis and cirrhosis are also more prevalent in those with T2DM.[55] The liver contributes to the insulin resistance seen in T2DM, and hepatic insulin resistance is a key factor in NAFLD.[76] Insulin resistance promotes the accumulation of triglycerides in the liver, leading to steatosis. Many studies have investigated the causal direction of the association between NAFLD and T2DM, and found NAFLD is a risk factor for T2DM; patients with NAFLD are five times more likely to develop T2DM.[77,78]

Dyslipidaemia is a condition characterised by abnormal amounts of lipids in the blood, and often accompanies T2DM and obesity.[79] This can be a risk factor for NAFLD independently, as circulating triglycerides and cholesterol are taken up by the liver and are stored, causing steatosis.[80]

The risk factors outlined above are the main phenotypic risk factors for NAFLD, but a number of others also affect NAFLD development. These include hypothyroidism, sleep apnoea, hypopituitarism and polycystic ovary syndrome.[81,82] Often individuals have several of the discussed risk factors, contributing various amounts to NAFLD risk.

1.6 NAFLD Genetic Background

NAFLD is complex and multifactorial, and as a result there are many genetic influences affecting different pathways associated with the disease. The first risk variant discovered by GWAS and most widely replicated result is *PNPLA3* rs738409, discovered in 2008.[52] This variant increases risk of simple steatosis as well as NASH and fibrosis.[83] The protein encoded by *PNPLA3* is adiponutrin, which plays an important part in the breakdown of triglycerides in

the liver (lipolysis). As a consequence of this downregulation of lipolysis, triglycerides build up to high levels causing steatosis. Further to the increased steatosis, *PNPLA3* rs738409 is also associated with inflammation and fibrosis, through regulation of NF-κB.[84] This *PNPLA3* variant is common, with a MAF of ~0.20 in Europeans and ~0.50 in Hispanics.[85] It has been the focus of a number of studies as a drug target but so far without success.[86]

Subsequently, variants affecting all stages of NAFLD have been discovered through GWAS and candidate gene studies. [87] Notable genes are shown in table 1 below, with their effect and the pathway which they act on.

Table 1 Genetic Variants Associated with NAFLD

Gene	SNPID	Discovery	Primary Effect	Role
PNPLA3	rs738409	GWAS	Increased steatosis	Decreases lipolysis
TM6SF2	rs58542926	GWAS	Increased steatosis	Decreases VLDL secretion
GCKR	rs780094	GWAS	Increased steatosis	Increases hepatic glucose uptake
NCAN	rs2228603	GWAS	Increased steatosis, NASH and fibrosis	Association with *TM6SF2*
COL13A1	rs1227756	GWAS	Liver enzymes and NASH	Inflammatory response
MBOAT7	rs641738	GWAS	Steatosis and fibrosis	Reduced protein production, not fully understood

The majority of genetic variants known to influence NAFLD have been discovered via GWAS, though a number of effects have been found through candidate gene studies.[88] These include the variant *MTP* rs1800591, which alters the activity of lipid transporters increasing steatosis.[89]

Heritability estimates for NAFLD vary between 20% and 70%, with a large portion of genetic variance shared between NAFLD and fibrosis.[90] Despite several genes having significant effects, genetic risk scores for NAFLD have been inconsistent in their ability to predict NAFLD. Nobili et al. found a genetic risk score based on 4 SNPs predicted NASH better than a clinical model.[91] Other studies have found genetic models to perform worse than those based on clinical parameters, and some that adding genetic data does not improve accuracy of clinical models.[92,93]

As well as prediction, analysis of genetic influences of disease can aid understanding of the pathology of the condition. This in turn can aid drug discovery, as it can elucidate pathways which can be up or downregulated to treat the disease.[94] This method has been applied to NAFLD, with research investigating the efficacy of inhibiting *HSD17-B13* for example.[95]

1.7 Gaps in Understanding and Research

A large number of research questions regarding NAFLD remain unanswered in current literature and require further research. One of the key questions is why certain individuals get NAFLD and others do not. Another clinically important question is why certain individuals progress to more severe stages of NAFLD, whereas some do not. Phenotypic and genotypic risk factors for NAFLD are known there remains large amounts of thus far unexplained variance in the previous two research questions. Factors such as diet, exercise, smoking, lifestyle and socioeconomic status have been investigated and found to have significant

effects on NAFLD, and must be taken into account when describing a full model of NAFLD risk.[4,96]

Much of the pathogenesis of NAFLD has been studied, and mechanisms of steatosis, NASH and fibrosis have been explained.[97] There are numerous pathways involved however, and a complete model of NAFLD has not been illustrated. Analysis of correlates of NAFLD, both phenotypic and genotypic, are both common methods of elucidating factors which affect NAFLD.

NAFLD is challenging to diagnose, as it lacks outward signs and symptoms in most cases.[25] This means that NAFLD is underdiagnosed in most clinical settings. This thesis will investigate NAFLD diagnosis methods and apply a definition to the GoDARTS, SHARE and Tayside and Fife cohorts from Scotland, as well as the DMDSC cohort from Chennai, India. This definition will be validated against other known NAFLD definitions and correlates. An investigation into the morbidity and mortality of patients with NAFLD will be conducted, with particular emphasis on analysing causes of death. Recent studies have found NAFLD to be associated with a range of extrahepatic conditions, including cancer.[58,59] This thesis will investigate whether NAFLD has any association with cancer incidence, as well as the effect of this on the shorter lifespans seen in those with NAFLD.

GWAS methods will be used to investigate genetic modifiers of NAFLD risk in the Scottish and Indian cohorts. Studies of the genetics of NAFLD in individuals of European descent are plentiful in the literature, though not specifically of Scottish cohorts.[98] Genetic studies of NAFLD in India have been published, findings significant effects of many known NAFLD risk variants.[99,100] There have been no reported GWAS studies of NAFLD in Indian individuals. This thesis will run GWAS analyses in Scottish and Indian cohorts, and compare the effects of different NAFLD risk variants in each population.

Co-agonists for GLP1R and GCGR have been deployed with some success for the treatment of diabetes, and have been investigated as a NAFLD therapy also.[101–103] A number of variants in the *GCGR* and *GLP1R* genes have significant effects on their namesake receptors, altering clinically relevant parameters to NAFLD, such as blood sugar and response to anti-diabetic drugs .[104,105] Effects of variants in these genes have never been shown in for NAFLD. This thesis will investigate the role of variants in these genes in NAFLD, and interactions between variants in *GLP1R* and *GCGR*. If variants with significant effects are found, this data may be effective in the application of personalised medicine to the deployment of GLP1R/GCGR co-agonists for the treatment of NAFLD.

Endothelin receptor antagonists (ETRAs) have been used successfully for the treatment of pulmonary arterial hypertension (PAH).[106–108] A number of pathological overlaps between PAH and NAFLD mean that ETRAs have been investigated as a potential treatment for NAFLD. It is thought that the downregulation of the endothelin receptors will reduce activity of the hepatic stellate cells (HSCs) which are primarily responsible for the production extracellular matrix (ECM) which characterises fibrosis.[109–111] There are a number of common genetic variants in *PHACTR1, EDN1, EDNRA* and *EDNRB* which have effects on endothelin and endothelin receptor activity.[112–115] No studies have previously reported associations between variants in these genes, and NAFLD. The current thesis will explore these genes and their associations with NASH, fibrosis and a number of related phenotypes. Showing effects of these genes on NASH and fibrosis would demonstrate clear effects of endothelin on the development of the conditions, and potentially allow for personalised medicine in those treated with ETRAs for NAFLD.

1.8 Aims of the Current Study

The aims of the current project are as follows:

-Develop a NAFLD definition and apply it to the GoDARTS, SHARE and Tayside and Fife Diabetics cohorts in Scotland, and the DMDSC cohort in India.

-Describe NAFLD and related phenotypes, and investigate morbidities and mortalities associated with NAFLD in Scottish cohorts.

-Analyse genetic determinants of NAFLD in Scottish and Indian cohorts using GWAS methods.

-Investigate the effects of specific candidate gene SNPs (related to GLP1R and endothelin) on NAFLD and associated phenotypes in Scottish and Indian phenotypes.

2 Development of a NAFLD Definition

2.1 Abstract

NAFLD is a chronic disease which often goes unnoticed due to the challenges of diagnosis. This is due to the lack of apparent signs and symptoms in most patients. A number of diagnosis methods exist with varying practicality and accuracy. The aim of this chapter is to develop and validate an accurate technique of defining NAFLD in large retrospective cohorts.

NAFLD diagnosis techniques from previous literature are considered and described. A NAFLD diagnosis technique based on elevated alanine transaminase (ALT) levels is used, as ALT is a commonly measured marker of liver damage. This definition is applied to the GoDARTS, SHARE and Tayside and Fife Diabetics (T&F) cohorts. A similar definition based on a single ALT level is applied to the DMDSC cohort.

A NAFLD definition based on two raised ALT levels measured at least months apart was applied, and had good sensitivity in GoDARTS. (97.4%) This definition was also accurate in SHARE and T&F, with sensitivities of 75.3% and 94.6% respectively. Thresholds of 19U/L and 30U/L (for women and men respectively) were used as upper limits of normal, based on recommendations of previous literature. An interval of at least 3 months between raised measurements was applied to reduce the likelihood of misclassifying acute liver injuries as NAFLD. This measure is practical, as most patients in GoDARTS have had a number of ALT measured. (Mean = 32.5 measurements each)

An accurate and practical NAFLD diagnosis method is developed and validated in this chapter. This novel definition may be useful for defining NAFLD in future studies using large retrospective cohorts, and analysing the genetics and epidemiology of NAFLD in these. This also may serve as a useful clinical risk estimating tool.

2.2 Introduction

The diagnosis of NAFLD can be challenging, as the condition is usually asymptomatic, progresses slowly and often has a benign clinical course.[116,117] This chapter covers the development of a NAFLD case definition based on raised alanine transaminase levels, in three large retrospective cohorts which had data for electronic health records (EHRs). The aim of this study is to develop a sensitive, accurate and functional NAFLD phenotype which can be used clinically and applied in research utilising retrospective cohorts

NAFLD usually lacks outwardly visible symptoms or signs, making it challenging to diagnose.[118] NAFLD can cause fatigue and pain in the upper right abdomen, but is usually completely symptomless.[119] More advanced liver disease in NAFLD, such as cirrhosis, can be accompanied by visible signs such as ascites and jaundice.[120] These symptoms usually occur in advanced liver disease, which only a minority of NAFLD patients progress to.[4] Diagnosis of the condition long before this happens is desirable as early intervention can improve disease outcomes.[121]

Many diseases don't present immediately obvious symptoms, and require clinical investigation to diagnose. These often have a reliable marker that can be used to infer the presence of the disease. An example of this is type 2 diabetes (T2DM). HbA1c is commonly used to diagnose T2DM, as it is an indicator of blood glucose levels over the last 2-3 months.[122] HbA1c is a simple blood test, which can be performed routinely and is cost effective.[123] Although there are a number of methods and tests for NAFLD, currently there is not a perfect, universally accepted and practical method for the diagnosis of NAFLD in real world cohorts.[124]

One of the main motivations for diagnosis of disease is to allow treatment. This is somewhat more complex in NAFLD, as there is currently no specific pharmacological intervention

recommended for NAFLD.[125] Despite this, there is evidence that NAFLD is reversible, even at the stage of cirrhosis in some cases.[126] Lifestyle modification has been the standard recommendation for NAFLD treatment for some time, and is included in NICE guidelines.[44] Weight loss and physical exercise have both been shown to reduce steatosis severity independently of each other. [68,127]

Diagnosing NAFLD is useful for research purposes.[128] An accurate NAFLD phenotype may be key factor in the epidemiological and genetic studies that seek to characterise the disease. These studies may aid drug development, pathways highlighted as important in the pathogenesis of NAFLD can be targeted. For NAFLD clinical trials, a NAFLD phenotype which is sensitive and specific is important in assessing the efficacy of potential treatments. These factors mean that despite the lack of a specific pharmacological treatment for NAFLD, there is great scientific and medical interest in developing methods to diagnose NAFLD.[129]

A key epidemiological phenomenon of NAFLD is underdiagnosis in most populations.[25] There are a number of estimates of global NAFLD prevalence, but the prevalence of diagnosed NAFLD lags far behind even conservative estimates. Alexander et al. found that the prevalence of NAFLD diagnoses was 1.85% in analysis of four large European cohorts. This is in contrast to the estimate of 20% NAFLD prevalence which Alexander et al. suggest, which is modest compared to the estimates of other studies such as Younossi et al. provide a global estimate of 25% prevalence.[49] This is higher in many regions, as Latino individuals tend to experience younger onset of NASH compared with Caucasian individuals for example.[130]

The current chapter considers existing literature on the diagnosis of NAFLD with the use of a number of methods. The advantages and disadvantages of each method are compared with respect to accuracy and practicality as a method of diagnosing NAFLD in large retrospective

cohort studies. A NAFLD case definition based on elevated ALT levels was developed and validated in the GoDARTS, SHARE, and T&F cohorts. ALT was selected as it is a specific marker of liver damage, and is routinely measured, making it an effective method of NAFLD diagnosis in large retrospective cohort studies.[131]

2.3 Methods of NAFLD Diagnosis

A number methods of diagnosing NAFLD have been developed and used, and generally there is a trade-off between the accuracy and practicality of measures. The following section discusses NAFLD diagnosis methods with their merits and disadvantages.

2.3.1 Biopsy

The gold standard method of NAFLD diagnosis is liver biopsy.[132] A small core tissue sample of the liver is taken with a needle which is then evaluated by a pathologist.[133] This method is accurate for detecting the presence of steatosis in hepatocytes and for diagnosing later stages of NAFLD such as fibrosis and cirrhosis. Liver biopsies are invasive procedures, and can lead to serious complications. These include haemorrhage, pneumothorax, biliary peritonitis, and even death.[134] As well as the risks of biopsy, inter-clinician variability in evaluation can influence the accuracy of diagnosis. [14] Steatosis, inflammation and fibrosis can often present unevenly throughout the liver, with some areas unaffected while others show severe NAFLD progression. This can lead to inaccuracy and false negative results. These factors combined with the high prevalence of NAFLD mean that biopsy is an impractical method of diagnosis, and often the risk is not worth the reward. Biopsy can be a useful tool for patients who are considered at greater risk of advanced stages of NAFLD such as fibrosis and cirrhosis, but non-invasive methods are more appropriate for NAFLD diagnosis.[125]

2.3.2 Imaging

Imaging methods such as ultrasound and MRI can be used for NAFLD diagnosis.[135] These methods can be accurate in detecting the presence of steatosis in the liver, as well as fibrosis and cirrhosis.[136] The primary advantage of imaging methods over biopsy is that they are non-invasive and do not pose the same risk of complications that biopsy does. Eddowes et al. found MRI was more cost effective than biopsy for diagnosing NAFLD. In two cohorts from NHS hospitals in the UK, the study showed MRI could save £150,218 per 1000 patients.[137] Ultrasound scanning techniques can be used at a patient's bedside in clinical settings, giving immediate results.[138]

Ultrasound is used effectively to grade fatty liver disease, and is more cost effective than MRI.[139] Ultrasound has been shown to accurately detect moderate-severe NAFLD from no NAFLD present.[140] As well as a binary diagnosis of NAFLD or not NAFLD, the severity of steatosis can be assessed by ultrasound. Clinicians often grade NAFLD on a 4 point scale: normal, intermediate, moderate and severe. Ultrasound techniques are not as accurate as biopsy, but provide more information about the current NAFLD status than a binary outcome.[141]

Similarly to biopsy however, Hamer et al. found the accuracy of imaging methods can be affected by inter-clinician variability.[142] Accuracy may also be affected by variation in equipment, equipment settings and the specific imaging modality used. Imaging methods may not be as accurate as biopsy, nor provide as much information about the status of liver disease in some instances. For example, Saadeh et al. found that there was no reliable method of discerning NAFLD from NASH using any radiological technique.[143] Further to the issues of accuracy, imaging techniques may be impractical due to resource limitations and time consumption. Patients may experience waiting times of several weeks for diagnostic procedures which are in high demand within health services.[144] Imaging processes are not

usually available at GP surgeries which means they are not practical for NAFLD diagnosis in the general population.

2.3.3 Blood Biomarkers

Blood based biomarkers are routinely used for quick and accurate diagnosis in many diseases including NAFLD.[31] These are non-invasive, rapid and usually economical when compared with biopsy or imaging.[145] Another advantage over biopsy or imaging is that blood can be drawn in general practice surgeries or in the community, making this form of testing for NAFLD very accessible. Many patients have blood taken routinely, allowing consistent monitoring of biomarkers. Multiple biomarkers can be tested at once and results combined to make disease scores, as considered below. Common biomarkers used for NAFLD and more advanced stages thereof include alanine transaminase (ALT), aspartate transaminase (AST), gamma-glutamyltransferase (GGT), ferritin and platelets.[146] These are considered liver function tests. (LFTs) Among these LFTs, ALT is a commonly used biomarker in NAFLD research and clinical practice.[131]

2.3.4 Scores and Indices

Scores and indices to estimate the likelihood and severity of NAFLD have been developed. These use a variety of often biochemical and anthropometric measures, and combine them in some numerical formula. An example of this is the NAFLD activity score (NAS).[147] This uses a combination of observed steatosis, lobular inflammation and liver damage to diagnose NAFLD and NASH. This is impractical as a method for diagnosis of NAFLD in the general population however, due to the pitfalls of biopsy discussed above.

FLI

A non-invasive scoring system for NAFLD is the Fatty Liver Index (FLI).[148] This combines BMI, waist circumference, GGT and triglycerides using the formula in figure 2-1 below.

$$\mathrm{FLI} = \frac{e^{0.953*\log e\,(\text{triglycerides}) + 0.139*\mathrm{BMI} + 0.718*\log e\,(\mathrm{ggt}) + 0.053*\text{waist circumference} - 15.745}}{1 + e^{0.953*\log e\,(\text{triglycerides}) + 0.139*\mathrm{BMI} + 0.718*\log e\,(\mathrm{ggt}) + 0.053*\text{waist circumference} - 15.745)} * 100}$$

Figure 2-1 Fatty Liver Index calculation formula

Bedogni et al. found that the FLI could detect NAFLD with an accuracy of 0.84. They found that ALT was predictive of NAFLD, but GGT performed better in their model. Subsequent studies have validated FLI against other NAFLD diagnosis methods such as imaging techniques, and found it performs well.[149]

This can be applied in the GoDARTS cohort where 4,164 eligible patients had the required data to calculate their FLI score. This represents about 30% of the eligible GoDARTS population, so is useful for replication and verification of results but the missingness of GGT measurements make FLI less useful for NAFLD diagnosis than definitions using more commonly taken measures.

Hepatic Steatosis Index

Another score for NAFLD diagnosis is the Hepatic Steatosis Index (HSI).[150] This uses AST(U/L), ALT(U/L), BMI(kg/m^2), sex and diabetes in the formula below to rule in or out NAFLD.

$$\mathrm{HSI} = 8 \times \frac{\mathrm{ALT}}{\mathrm{AST}} + \mathrm{BMI}\ (+2\ \text{if female}, +2\ \text{if T2DM present})$$

Figure 2-2 Hepatic Steatosis Index calculation formula

This index can also be applied to a subset of GoDARTS where 1,493 patients had the requisite data measured before sign up to the study. Like the FLI, this is useful for verification of results in GoDARTS and SHARE, but the number of patients without AST levels measured make this impractical as a main NAFLD definition.

APRI

The APRI (AST to Platelet Ratio Index) score is also used to diagnose hepatic fibrosis.[151,152] APRI is calculated as follows.

$$\text{APRI} = \frac{\text{Aspartate Transaminase(U/L) / Aspartate Transaminase Upper Limit of Normal (U/L)}}{\text{Platelet Count } (10^{-9}/\text{L})} \text{ x100}$$

Figure 2-3APRI calculation formula

The upper limit for normal AST is taken as 40U/L in most studies.[151] This index has been found to correlate strongly with fibrosis and cirrhosis, but has little predictive value for steatosis however.[153]

The FLI, HIS and APRI indices are useful markers for steatosis and fibrosis, but the lack of availability of certain data, especially AST measurements, make these impractical as a main technique for defining NAFLD for the analysis in the current study, as the analytical requirements of these measures would result in the exclusion of a significant portion of the GoDARTs, SHARE and T&F cohorts.

2.3.5 Summary of Techniques

The previous section covers a number of techniques used for the diagnosis of NAFLD in clinical and research settings. Based on evidence from previous literature, and the availability of relevant data, a NAFLD definition based on elevated ALT levels is selected for the current

study. The following section describes the association between NAFLD and ALT, covers the development of a NAFLD definition based on ALT levels.

2.4 ALT Based NAFLD Definition

2.4.1 Biology

ALT is part of the gluconeogenesis process and plays a major role in metabolic homeostasis. It is an enzyme which catalyses the transfer of α-amino acids from alanine to create pyruvic acid.[154] This process occurs in the liver, and thus ALT is found in its highest concentrations in the liver and in low concentrations elsewhere in the body.[155] ALT is found in the cytoplasm of hepatocytes, and when hepatocellular injury occurs, ALT is released into the bloodstream. As a result, ALT levels are commonly used biomarker of liver damage.[156] The mean half-life of serum ALT is 47 hours, meaning that measured ALT levels can be affected by acute events.[154]

Elevated ALT levels have been observed in those with NAFLD in numerous studies. [31,157–160] ALT levels are known to correlate with liver fat percentage; Phillips et al. estimate this correlation to have R^2= 0.51 for example.[161] Other studies have found ALT to be a useful predictor of NAFLD in those with steatosis confirmed by MRI.[162] Maximos et al. found that ALT levels were a strong predictor of hepatic triglyceride content.[163] A number of previous epidemiological studies have used ALT levels to define their NAFLD cases. Wong et al. investigated the showed the link between T2DM and NAFLD with ALT as a surrogate marker.[164] Cross sectional case-control studies such as Yoo et al. in 2008 used ALT to define paediatric NAFLD.[165] ALT has been used as a surrogate marker to track resolution or progression of NAFLD also. [166,167]

2.4.2 Practicality

An ALT based NAFLD definition is a pragmatic choice for defining NAFLD in the cohorts used in this project for a number of reasons. A key factor which makes ALT a useful biomarker for NAFLD is that ALT levels are commonly measured in both GP and hospital settings. The majority of patients in GoDARTS and SHARE have ALT measurements taken, and in many individuals a large number of measurements have been taken over many years. In GoDARTS for example, patients had a mean number of 32.5 ALT measurements each. This allows the assessment of NAFLD over a longitudinal period with some degree of regularity, and an interesting analysis of ALT levels over time compared with other clinical outcomes. This reflects what is seen in primary care settings in general, as ALT levels are monitored as a key indicator of NAFLD.[42]

2.4.3 Specificity versus other diseases

Serum ALT levels are specific markers of liver damage due to their high concentrations in the liver under normal circumstances. ALT is a general marker of hepatocellular injury however, and is not specific to NAFLD.[156] There are a variety of causes of liver damage which result in raised ALT levels; therefore to use ALT for NAFLD diagnosis, all other causes of liver injury must be ruled out.

One of the major alternative causes of abnormal ALT levels is alcoholic liver disease (ALD).[168] Excessive alcohol consumption can cause steatosis of the liver much like NAFLD.[169] The demarcation between ALD and NAFLD is not clear in a large portion of patients, as the two diseases have similar features, and can lead to fibrosis, cirrhosis and liver cancer. Patients with fatty liver disease may have their condition attributable to both alcohol and metabolic aetiologies.[33] The generally accepted upper alcohol consumption limits for NAFLD are <20 grams per day for women and <30 grams per day for men.[42] It is likely that in patients with fatty liver who drink alcohol but below these limits, alcohol may still

contribute to the disease. Previous studies have shown that ALT levels are more correlated with the non-alcoholic form of fatty liver disease, and AST tends to be higher in ALD.[170]

Other liver insults which can raise ALT levels include viral and immunological hepatitis. These are usually diagnosed with blood tests, the results of which are reported in immunology and virology files in EHRs. Patients with positive serological tests for anti-smooth muscle antibody, antinuclear antibodies or anti-mitochondrial antibodies, or any positive serology for hepatitis B surface antigen or hepatitis C antibody, or mention of cause of liver disease in medical records are excluded from the study.

Aside from other liver disease, factors which do not warrant exclusion can affect ALT levels. There is evidence that certain drugs at therapeutic doses, including paracetamol, can raise ALT levels in the absence of any clinically evident liver damage.[171]

2.4.4 Sensitivity for NAFLD

Despite the large volume of evidence that NAFLD causes elevations in ALT levels, there have been a number of studies that report the complete spectrum of NAFLD in patients with normal ALT levels.[172] Mofrad et al. found that some patients within the normal ALT range, 50U/L for women and 75U/L for men, had steatosis, fibrosis and cirrhosis. These thresholds are however are considerably higher than those adopted by most modern studies, and may not reflect truly "normal" ALT values. Sanyal et al. had similar results, but they too used a high ALT threshold, at 40U/L.[173] Despite higher thresholds, both studies identified individuals with low ALT < 15, with NAFLD. Both of these studies used cross-sectional techniques, with a single ALT measurement on the date of the study and this is a potential source of error as follows. In severe cases of cirrhosis, ALT levels can decrease as the number of hepatocytes diminishes and less ALT is released into the bloodstream.[174] Lominadze et al, found many patients with cirrhosis had normal ALT levels.[175] In GoDARTs there are patients with normal

ALT levels at the time of cirrhosis diagnosis, but had previously had raised ALT levels. ALT levels have also been found to vary by age, following an inverted U shape over a lifetime.[176] The retrospective application of a NAFLD definition based on ALT levels allows an "ever raised" case definition, increasing sensitivity compared with cross-sectional techniques.

The sensitivity of ALT as a diagnostic marker for NAFLD has been found to differ between those with and without T2DM.[177] Kotronen at el. found that individuals with T2DM have 80% more liver fat than comparable non-diabetic individuals, but using ALT as a biomarker underestimates the prevalence of NAFLD in those with T2DM. Indeed, diabetic patients with comparable ALT levels had between 40% and 200% higher liver fat content than their matched non-diabetic counterparts. Given the high proportion of individuals with T2DM in GoDARTS, this may mean that an ALT based NAFLD definition will underestimate NAFLD in the cohort.

2.4.5 Thresholds

For diagnosing disease with ALT levels, a threshold between healthy and unhealthy is needed. There have been many thresholds used clinically and in studies. The most commonly used threshold is 40U/L, which was first set in the 1950's and is still commonly used.[178] There has been a trend towards lowering these limits. Studies in the 1990's saw the limits of 30U/L to 50U/L on average.[178]

It is believed that many of the cohorts used to define abnormal ALT levels contained individuals with undiagnosed liver insults, leading to high ALT levels being used for thresholds of the healthy range.[179] In 2002, Prati et al. suggested lowering these even further to 19U/L for women specifically, based on the results of their study looking at 6,835 individuals without any viral liver insult.[180] Using these values, Kunde et al. investigated the effect on the rate and accuracy of NAFLD diagnosis.[179] They found that sensitivity was

improved, at the cost of decreased specificity. Tomizawa et al. found that using the threshold of 19U/L for both male and female patients was a useful marker for screening for NAFLD.[157]

The thresholds suggested by Prati et al. (30U/L for men, 19U/L for women) were used for the current study. These were chosen as there is good evidence that these thresholds are the true ULNs for ALT in healthy individuals, and that these thresholds are useful in screening for and predicting NAFLD.

2.4.6 Temporal Variation in ALT

ALT levels can vary over time, therefore the time at which blood is drawn for ALT tests can affect the result. This can embed another source of variance in ALT levels which cannot be accounted for. The time of day at which a blood sample is drawn may affect the ALT measurement. In mice it has been found that liver function oscillates daily in response to circadian rhythms and ALT levels fluctuate throughout the day.[181] Given the 47 hour half-life of serum ALT, this fluctuation is represents a delayed representation of ALT secretion into the bloodstream. There is some evidence that ALT levels vary by time of day in humans as well as mice, and that they are highest in the latter half of the day.[182,183] Factors mentioned above, such as paracetamol intake or low level alcohol intake, can vary which may cause variation in ALT levels.

Sattar et al. found that sustained increases in ALT levels were predictive of progression to T2DM, whereas isolated increases in ALT were not predictive.[184] Given the link between NAFLD and development of ALT, the ALT level increases appear to be good markers of NAFLD in this study.[185] Though this study looked at changes in ALT levels and the current study uses absolute values, this supports the notion of sustained ALT elevation as a useful diagnostic marker for NAFLD.

As mentioned above, cross-sectional techniques may underestimate the presence of NAFLD if based on ALT levels alone. Mofrad et al. showed all stages of NAFLD in a number of patients with healthy ALTs.[172] In GoDARTs, there are a large number of patients who experience raised ALT levels followed by normal ALT levels. In fact, there is evidence that ALT levels decrease in the later stages of fibrosis.[174] This could indicate that patients with severe fibrosis and normal ALT have previously had raised ALT levels. The sensitivity of the NAFLD definition in this study is improved by using longitudinal data, as raised ALT levels at any point in time are used to form the NAFLD phenotype.

When making diagnoses based on biomarker levels, using a sequence of measurements taken at different times can be useful in increasing accuracy. This technique is established for diagnosis other diseases; for example chronic kidney disease, where changes in glomerular filtration rate and urine output over time are used for diagnosis and staging.[186] Using several ALT measurements to define NAFLD can improve specificity, as acute cases of liver damage with raised ALT levels in a short period of time are not categorised as NAFLD events. The current study uses any two raised measurements that are a minimum of three months apart to define NAFLD. This means that single incidents of raised ALT levels are not counted as NAFLD, increasing specificity. The minimum interval of 3 months between raised ALT measurements makes this chronically raised ALT levels, consistent with the chronic nature of NAFLD cases.

2.5 NAFLD Definition in the Current Study

The ALT based NAFLD phenotype for this project is defined as follows; 2 or more raised ALT measurements in the absence of alternative cause of liver disease, at least 3 months apart. In contrast to previous literature, this adds a higher degree of certainty that the condition is chronic. The retrospective application of this definition allows increased sensitivity for NAFLD diagnosis, as discussed in the sections above.

The thresholds of 30U/L for men and 19U/L for women for the upper limit of normal ALT were chosen, based on recommendations by Prati et al.[180]

2.6 Application in GoDARTS, SHARE, and Tayside and Fife

2.6.1 Introduction to cohorts

Three large cohorts were available for analysis in this project. These comprised of patients within the Tayside and Fife NHS health boards.

The first cohort used was GoDARTS.[187] This was a case-control type 2 diabetes study based in Tayside, Scotland. This cohort comprised of electronic health records (EHRs) from 18,306 individuals, 10,149 of whom have T2DM. On patients' date of sign up they were phenotyped by biochemical and haematological investigations, anthropometric measurements and lifestyle questionnaires.

The second cohort used was SHARE, a cohort of 73,024 individuals who volunteered to allow their medical records to be used for scientific research, and is open to anyone in Scotland over the age of 16.[188]

The third cohort includes patients from the East of Scotland from population-level data on individuals with type 2 diabetes across the regions of Tayside & Fife (T&F) in Scotland. Clinical data is made available through the Scottish Care Information-Diabetes Collaboration (SCI-DC) system. This cohort comprise 89,553 individuals at the time of study.

2.6.2 Data available: ALT, other necessary data such as immunology and virology

Medical records from the NHS Tayside and Fife boards are available for patients in each cohort. These extend back as far as the beginning of 1987 in some cases, when records were digitised. These include admissions, deaths, prescriptions, biochemistry, demography and a number of other files for each patient in the cohort.

For the GoDARTS cohort, alongside the NHS EHR records, additional data was collected on the date of sign up. These records included data for biochemical measurements such as triglycerides, cholesterol and HbA1c, anthropometric measures such as BMI, Waist and Weight and lifestyle factors including smoking, drinking and exercise.

2.6.3 Exclusions

The exclusion of patients with other forms of liver disease which may affect the accuracy of NAFLD diagnosis is performed using data from EHRs. For each cohort a smoking and alcohol file is available for diabetic patients. This includes a column with patients mean alcohol intake per week. Men who drink more than 30g per day on average and women who drink over 20g per day on average are excluded. Patients are also excluded if they have any alcohol excess related condition reported in their medical records. In GoDARTS, 1,175 individuals are excluded due to alcohol consumption or alcohol related hospital admission.

Patients are excluded from analysis if they have features of other chronic liver disease, including: any positive serological tests for anti-smooth muscle antibody, antinuclear antibodies or anti-mitochondrial antibodies, any positive serology for hepatitis B surface antigen or hepatitis C antibody, or mention of cause of liver disease in medical records. In GoDARTS, 1,200 individuals are excluded do to alternate causes of liver disease reported in EHRs.

2.6.4 Prevalence rates in each cohort

The lifetime prevalence rates were calculated for each cohort using all the data available for each patient, after patients with alternative causes of liver disease were excluded from analysis. This is calculated as the presence of NAFLD any time between first available medical record and last medical record, either due to death or the end of the study follow up period.

The lifetime NAFLD prevalence rates as assessed using the ALT based NAFLD definition outlined above are as follows: GoDARTS: 68.24%, SHARE: 49.26%, T&F: 64.09%.

2.6.5 Specificity and sensitivity versus ICD10 codes, and NASH

The NAFLD diagnosed by ALT were analysed compared to ICD10 codes recorded in EHRs'. The sensitivity of this definition was 97.4%, and the specificity was 32.0%. Using the same method, in SHARE this definition has a sensitivity of 75.3% and specificity of 54.2%, and in T&F the definition has a sensitivity of 94.6% and specificity of 38.3%. Due to the underdiagnosis of NAFLD in clinical settings, the specificity of the definition is of little utility or importance for validating NAFLD in the current study.[25]

To further validate this phenotype, positive control tests are run against chronic kidney disease in GoDARTS, as this has been shown to associate with NAFLD.[189] NAFLD was found to associate with increased incidence of chronic kidney disease in a Cox proportional hazards model adjusted for sex, T2DM, age, and BMI. (HR = 1.32(1.25 – 1.39), $p = 4.8 \times 10^{-5}$)

A positive control test with the well-known NAFLD risk variant *PNPLA3* rs738409 was conducted in the GoDARTS cohort.[190] In a logistic regression with an additive model, adjusted for age and sex, PNPLA3 rs738409 was associated with increased NAFLD at the beginning of the study. ($p = 1.09 \times 10^{-8}$, OR = 1.32(1.24-1.41)) This result is similar to other results found for the effect of *PNPLA3* on NAFLD, such as Wang et al. with an OR of 1.52.[191]

2.7 Application in GoDARTS Cohort

The NAFLD definition used in this project was decided upon after consultation of the current literature, and analysis of sensitivity and specificity versus NAFLD recorded in medical records, as diagnosed by physicians. Several other NAFLD definitions are considered in the sections above, using different ALT thresholds, and different temporal rules. The current

section outlines some commonly used or suggested alternative NAFLD diagnosis methods, and how these definitions performed in GoDARTS.

2.7.1 Thresholds

A number of thresholds for ALT were trialled. These are as follows:

- 30U/L for men, 19U/L for women – Prati et al. [180]
- 30U/L for men, 30U/L for women – Kunde et al. [179]
- 40U/L for men, 35U/L for women – Neuschwander-Tetri et al. [192]
- 25U/L for men, 17U/L for women – Miyake et al. [31]

2.7.2 Number and Timings of Measurements

A single case of a raised ALT is often used to diagnose liver disease, but as discussed above, using multiple measurements can improve accuracy. The following methods of using raised ALTs to define NAFLD were tested.

- 1 raised ever
- 2 raised ever
- 2 raised at least 3 months apart

The definitions of NAFLD listed above were compared using a number of metrics. This is shown in the table below. This table displays the prevalence rate of NAFLD with each definition, stratified by T2DM. It also shows the sensitivity and specificity of the definition when compared to NAFLD diagnoses in EHRs coded by ICD10 codes. The sum of specificity and sensitivity for each NAFLD definition is included in table 1. Finally, the table shows the difference between expected NAFLD rate and actual rate, stratified by T2DM. The expected rates in T2DM and non-T2DM were considered to be 70% and 30% respectively. [193] This calculation shows how much each definition is likely to over or underdiagnose NAFLD. This gives a notion of how credible each definition is, where a large difference

would suggest a poor definition of NAFLD. The results of this comparison are shown in the table below, with the main NAFLD definition for this project of two raised ALT measurements a minimum of three months apart highlighted in grey.

Table 2 - Comparison of Performance of NAFLD Diagnosis Methods in GoDARTS

Method	Male ALTs	Female ALTs	Prevalence Rate	Sensitivity	Specificity	Specificity plus Sensitivity	Diabetic NAFLD Rate	Healthy NAFLD Rate	Cumulative Difference from expected rate
1 raised	25	17	89.51	100.00	10.54	110.54	96.61	82.21	78.81
	30	19	83.91	100.00	16.17	116.17	93.01	74.54	67.55
	30	30	72.53	98.55	27.60	126.15	85.78	58.90	44.68
	40	35	58.52	98.55	41.68	140.23	72.86	43.76	16.62
2 raised outside 3 months	25	17	78.04	97.10	22.06	119.16	90.93	64.77	55.70
	30	19	68.24	97.40	31.90	127.55	83.01	53.02	36.04
	30	30	51.85	94.20	48.35	142.55	68.97	34.23	5.27
	40	35	35.81	85.51	64.43	149.94	50.77	20.42	28.81
2 raised ever	25	17	82.30	100.00	17.79	117.79	93.77	70.49	64.26
	30	19	73.90	98.55	26.21	124.76	87.54	59.89	47.43
	30	30	59.57	97.10	40.44	137.54	76.37	42.65	19.01
	40	35	45.28	92.75	54.95	147.70	60.87	29.24	9.89
2 raised in 1 year	25	17	72.83	98.55	27.30	125.85	89.29	55.89	45.18
	30	19	64.04	97.10	36.13	133.23	81.00	46.58	27.58
	30	30	50.01	94.20	50.14	144.34	67.20	32.46	5.26
	40	35	36.79	91.30	63.48	154.78	51.11	22.05	26.84

2.7.3 Comparison of Definitions

The comparison table above shows the performance of each of the definitions. The common trade-off between specificity and sensitivity evident in the results as there is a range of overall NAFLD rates as given by the definitions.

All definitions have fairly high sensitivity, ranging from 85.5% to 100%. Due to the relatively low number of cases of NAFLD reported in ICD10 codes, this test lacked the power to satisfactorily evaluate differences in sensitivity. However, consideration of ICD10 codes demonstrated that a high proportion patients with a NAFLD diagnoses in admissions data have had elevated ALT measurements at some point.

The specificity of the definitions is not as informative as the sensitivity, due to underdiagnosis of NAFLD in hospital admissions.[25] This means that NAFLD defined in medical records is specific but not sensitive compared to ground truth. Despite the low specificity of ALT based NAFLD definitions when comparing to NAFLD admissions, the ALT based definitions are likely closer to the actual NAFLD rate in the population.

Definitions based on more than one raised ALT measurement are more specific than one single raised ALT level. Using two raised ALTs in a year, or two outside three months both gave good sensitivity to NAFLD as phenotyped by hospital admissions.

The definition selected for this project (highlighted in the table) had good sensitivity but low specificity. This definition gave reasonable rates for NAFLD in diabetics and non-diabetics of 83.0% and 53.0% respectively. Although these are higher than rates estimated in current literature for diabetic and non-diabetic populations, the cohort has a high median age (73 years), which may account for high prevalence.[193]

Whilst some other ALT based NAFLD definitions that were tested have better performance metrics such as specificity, the definition for NAFLD of two raised ALTs separated by at

least three months was selected for a number of reasons. The thresholds of 30/19UL for men and women are chosen as the current literature suggests these are the true ULNs and two raised measurements at least three months apart as compared to cross-sectional techniques this increases specificity and diminishes risk of aberrant ALT fluctuation affecting the results. The measurements being at least three months apart ensures the definition is chronically raised ALTs, which adds accuracy as NAFLD is a chronic condition.

2.7.4 Patient Profiles, Morbidity and Mortality in NAFLD

The following section contrasts those with and without NAFLD, as defined by two raised ALT levels three months apart. Aside from differences in liver enzymes, individuals with NAFLD are known to differ from healthy individuals in a number of parameters. These are shown in the table 2 below, where the differences between groups for each variable are all statistically significant. ($p < 0.05$)

Table 3 - Patient Profiles Stratified by NAFLD Status in GoDARTS

	Healthy	SE	NAFLD	SE
N	6731		6950	
Diabetic	1662	24.7%	4605	66.3%
Female	2796	41.5%	3865	55.6%
BMI (kg/m^2)	27.2	4.8	30.9	6.1
Age (years)	61.9	13.7	64.9	11.5
Waist (cm)	99.3	11.4	107.1	12.7
	87.1	13.3	98.4	14.8
Weight (kg)	84.4	14.2	92.9	17.1
	70.2	14.4	79.6	18.3
Baseline ALT (U/L)	21.0	8.1	36.3	18.7
	16.1	12.9	27.1	19.3
Triglycerides (mmol/L)	1.6	1.0	2.1	1.4
HDL (mmol/L)	1.6	0.5	1.4	0.4

LDL (mmol/L)	2.8	1.0	2.4	0.9
Systolic Blood Pressure (mmHg)	137.5	20.1	140.5	19.1
Diastolic Blood Pressure (mmHg)	78.9	10.7	78.0	11.1
SIMD10	6.4	2.6	5.9	2.7

Those with NAFLD weighed more, are more likely to be diabetic, had higher BMI, are older, and had larger waists and higher triglycerides. These are known risk factors/covariates of NAFLD, and demonstrate the wider ranging metabolic correlates of NAFLD.[1,63]

2.7.5 Morbidity and NAFLD

NAFLD is associated with a range of morbidities. These include the range of NAFLD outcomes and complications; NASH, fibrosis, cirrhosis and hepatocellular carcinoma.[5] These more extreme endpoints are relatively uncommonly reported compared with NAFLD itself. NAFLD was associated with increased number of hospital admissions in GoDARTS, in a linear regression adjusted for sex, age, T2DM, BMI and SCSIMD10. ($\beta = 2.13(0.11 - 19.10)$, $p < 2x10^{-16}$) Although some of these admissions are due to liver conditions, NAFLD is still associated with more admissions when these visits are excluded. ($\beta = 2.11(0.11 - 19.34)$, $p < 2x10^{-16}$) This suggests that the majority of the increase in morbidity associated with NAFLD is not directly due to liver conditions.

Previous literature has linked NAFLD with a number of cardiovascular endpoints, and suggested this is the predominant cause of morbidity and mortality in NAFLD.[194] These include ischemic heart disease, myocardial infarction, and atrial fibrillation.[58] In GoDARTS, NAFLD was significantly associated with increased cardiovascular disease incidence (HR $=1.35(1.28 - 1.43)$, $p < 2x10^{-16}$) and with increased cardiovascular death. (HR $= 1.26(1.17 - 1.37$, $p = 1.55x10^{-8}$) Other conditions such as chronic kidney disease have been associated with NAFLD, which was also replicated in GoDARTS. [189]

2.7.6 Mortality and NAFLD

NAFLD has been linked to increased mortality by a number of studies.[5] Death caused directly by NAFLD, through failure of the liver due to cirrhosis or HCC, is the most common NAFLD related death cause investigated. These conditions are uncommonly reported in death certificates however. In GoDARTS for example, only 2.0% of individuals with NAFLD who have died have a NAFLD related condition listed in their causes of death. This is consistent with most findings about NAFLD mortality, which is dominated by cardiovascular and cancer death.[58,120]

This large difference between individuals with and without NAFLD cannot be explained fully by NAFLD specific conditions, and is due to other causes. Some studies have found associations between NAFLD and cardiovascular death for example.[120] To investigate this, individuals with NAFLD endpoints recorded in medical records are excluded from analysis. In GoDARTS, NAFLD (as diagnosed by ALT levels) still has a significant association with lowered death age. (β = -1.95(0.378 -5.15), p =2.74x10^{-7}) This demonstrates that there is clear increased mortality among those with NAFLD which is not caused by direct NAFLD endpoints. Though some of this may be attributable to the underreporting of NAFLD, it is likely that there is another cause or multiple other causes of this difference.

NAFLD has been associated with cardiovascular death, which is considered by some as the predominant cause of early death in NAFLD.[58,120] There has been some research into the link between NAFLD and extrahepatic cancers, which is the focus of the next chapter in this thesis. This investigates the link between NAFLD and cancer incidence, and cancer mortality and considers the role these play in the increased morbidity and mortality in NAFLD patients.

2.7.7 Genetics and NAFLD

NAFLD has a number of known genetic risk modifiers, including *PNPLA3* rs738409 which was analysed with the current NAFLD definition above.[52,90] This thesis investigates genetic determinants of NAFLD in subsequent chapters, beginning with a genome wide association study (GWAS). Candidate gene studies of variants related to GLP1R and endothelin are also conducted. The NAFLD definition developed in the current chapter provides an accurate and validated phenotype for these studies to produce valid results.

2.8 Application in DMDSC South Indian Cohort

Further to analysis in the GoDARTS, SHARE and T&F cohorts, access to a fourth cohort with clinical and genetic data was available. This cohort contained South Indian individuals, allowing for interesting comparison with Scottish cohort.

2.8.1 Introduction to cohort

The Indian cohort is from Dr. Mohan's Diabetes Specialities Centre (DMDSC) based in Chennai, India.[195] DMDSC is a large, privately-run chain of single-speciality hospitals and clinics for the treatment of diabetes and related comorbidities. This cohort comprises of 75,952 patients from Chennai and the surrounding area, all of whom have T2DM. The majority of patients are of South Indian origin.

2.8.2 Data Availability

The data from these patients is from a set of measurements taken on their first visit to the DMDSC clinic. During this visit, each patient had a series of demographic and phenotypic measurements taken. These included; HbA1c, BMI, ALT, HDL, age and sex.

Of the 75,952 unique patients with baseline biochemistry measurements, 33,194 of these patients had ALT measured at their first visit. This single ALT measurement is used to define

NAFLD in the DMDSC cohort. ALT measurements are available for patients at the date of their visit to the DMDSC centre. This allows a cross sectional analysis of the NAFLD prevalence in this cohort. The baseline measurements for most patients were taken within one year of their T2DM diagnosis date, allowing for interesting assessments of NAFLD in those recently diagnosed.

A number of these patients had genotypic data available for analysis giving 3,150 patients with genotype data and sufficient data for analysis of NAFLD defined by ALT levels, and 2015 patients had genotypic data and Fatty Liver Grade data.

2.8.3 Suitability of ALT as NAFLD Surrogate

The usefulness of ALT in this cohort is lessened due to lack of data providing other possible causes of liver disease. Due to the fact this data is from visits to a private clinic, routine healthcare similar to that of the GoDARTS and cohort was not available. This meant that it is not possible to exclude patients with alternative aetiologies for liver disease such as virological, immunological or alcohol related disease in the same manner as is done in GoDARTS. There is an alcohol consumption data field, with a binary yes/no whether the individual drinks alcohol. From this it is not possible to tell whether the individual drinks at levels which could cause ALD, therefore this variable is included as a covariate in the analysis.

The nature of the data also means it is not possible to use more than one ALT measurement for NAFLD diagnoses as there was no access to routine measurements. This meant that the NAFLD definition was based on one single ALT measurement at the time of enrolment.

2.8.4 Fatty Liver Grade in DMDSC

13,367 of patients in the DMDSC cohort have had abdominal ultrasounds administered. In the report, each patient was assigned a Fatty Liver Grade. (FLG) This was an index of hepatic

fatty infiltration scored in four ordinal categories. These were no fatty infiltration -0, mild fatty infiltration – 1, moderate fatty infiltration – 2 and severe fatty infiltration -3.[196] Ultrasound has been demonstrated as an accurate method of detection for steatosis of the liver.[157] Ultrasound is more specific for defining NAFLD than ALT levels, as liver fat itself is what is evaluated rather than liver enzymes. Where fatty infiltration is greater than 20% of hepatocytes, the sensitivity of ultrasound was found to be 90%.[197] Ultrasound techniques however are less effective in diagnosing the inflammation seen in NASH, as well as fibrosis.[198] Another caveat is that it is impossible to distinguish between NAFLD and ALD using this method, as they both cause steatosis and are histologically similar.[199]

Due to the superior performance of ultrasound for NAFLD diagnosis, and the lack of key data regarding alternative liver disease, ultrasound defined Fatty liver Score is the optimal method for NAFLD phenotyping in this cohort. As a result, the ALT measurements can be used as an adjunct, either as an additional phenotype or to aid validation of ALT based NAFLD definitions in other cohorts.

2.8.5 NAFLD Rate and Distribution of Fatty Liver Grade in the DMDSC Cohort

At the time of baseline measurement, the prevalence of NAFLD was 50.8%. The distribution of Fatty Liver Grades is shown in the table 3 below, with additional mean statistics of each group.

Table 4 - Comparison of Patients in DMDSC Cohort Stratified by Fatty Liver Grade

Fatty Liver Grade	0	1	2	3
Number (%)	2500 (18.7%)	4559 (34.1%)	5673 (42.4%)	635 (4.75%)
NAFLD Rate (ALT definition)	20.65%	25.93%	37.27%	41.87%

ALT (female/male)	25.24U/L / 30.70U/L	24.24U/L / 34.65U/L	32.19U/L / 45.19U/L	31.34U/L / 48.47U/L
BMI	24.30kg/m^2	25.90 kg/m^2	27.96 kg/m^2	30.75 kg/m^2
Sex (% Male)	58.98%	64.99%	68.16%	69.72%
AST	26.34U/L	26.69U/L	31.90U/L	32.17U/L

There was a strong association between FLG and NAFLD defined by ALT. (OR = 1.10(1.09 – 1.11), $p < 2x10^{-16}$) There were also significant associations between FLG and age, BMI, and AST. Higher FLG was seen in males.

2.9 Conclusion and Limitations

To conclude, the current section reviews previous literature on NAFLD diagnosis methods, and proposes a NAFLD definition based on chronically raised ALT levels. This definition was validated in three large Scottish cohorts, and has good sensitivity when compared to NAFLD diagnoses in EHRs. This chapter outlines a practical method for the diagnosis of NAFLD in the large retrospective datasets for research purposes.

This is an important finding, as diagnosing NAFLD remains an obstacle in NAFLD research and treatment. This methods uses data from retrospective medical records, thus requires no extra clinical measurements or investigations to take place making it cost effective and practical. Analysis revealed that patients with NAFLD were found to have greater risk of morbidity and mortality. This NAFLD definition is used in subsequent thesis chapters for the investigation of genetic modifiers of NAFLD risk.

This NAFLD definition may be useful in large scale datasets with data for thousands of individuals, but it likely results in a small number of false positives and false negatives. It has been shown that ALT levels correlate well with liver fat content, and there is evidence that the limits of normal ALT used in this study are accurate.[26,180] However, we are unable to

estimate the severity of the phenotype captured by our current NAFLD definition. This could mean that our NAFLD definition classifies those with liver fat <5% as NAFLD, overestimating the prevalence. The correlations with NASH as per medical records and other NAFLD indices suggests overall NAFLD is being captured reasonably well, but given the high prevalence positives cases for our definition in the cohorts, it likely overestimates slightly. This makes the definition of limited use on an individual level in clinical settings. The increased morbidity and mortality associated with this phenotype demonstrate the clinical relevance of the diagnosis method, though further work to stratify and identify those at risk may be more useful. The utility of diagnosing NAFLD in primary care settings is discussed in depth in the introduction to this thesis.

3 NAFLD and Cancer Incidence

3.1 Abstract

The aim of this study was to investigate the incidence of cancer in NAFLD patients and non-NAFLD controls, and the role of BMI in this relationship.

GoDARTS, SHARE and Tayside and Fife diabetics, three Scottish cohorts of 13,695, 62,438, and 16,312 patients respectively were analysed in this study. Participants in GoDARTS were a volunteer sample, with half having T2DM. SHARE were a volunteer sample. Tayside and Fife diabetics was a population level cohort. Patients with the relevant healthcare data available for analysis, and individuals with alternative causes of liver disease were excluded from the analysis.

NAFLD increased cancer incidence with a hazard ratio of 1.31 in a cox proportional hazards model adjusted for sex, type 2 diabetes, BMI, and smoking status (95% CI = 1.27 – 1.35, $p = 1.8x10^{-10}$). This was replicated in two further cohorts, and similar associations with cancer incidence were found for Fatty Liver Index, FIB-4 and NASH. Homozygous carriers of the common NAFLD risk variant *PNPLA3* rs738409 had increased risk of cancer. (HR = 1.27 (1.02-1.58), $p = 3.1x10^{-2}$) BMI was not independently associated with cancer incidence when NAFLD was included as a covariate. Finally, NAFLD was associated with increased risk of cancer death (HR = 1.40, 95% CI =1.33 - 1.47, $p = 3.7x10^{-6}$).

NAFLD is associated with increased risk of cancer incidence and death, as is *PNPLA3* rs738409, suggesting a causative relationship between NAFLD and cancer. NAFLD may be a major component of the relationship between obesity and cancer incidence.

3.2 Introduction

Non-alcoholic fatty liver disease (NAFLD) is the most common cause of liver disease globally, affecting around 25.2% of adults worldwide.[200] NAFLD, a spectrum of simple

steatosis to non-alcoholic steatohepatitis, is traditionally associated with endpoints which affect the liver, including fibrosis, cirrhosis and hepatocellular carcinoma (HCC).[201] The previous chapter in this these developed a definition of NAFLD and found increased numbers of hospital admissions and increased mortality in patients with NAFLD. The majority of these patients (>98%) had never had any NAFLD or even hepatological condition recorded in their admissions or death records. This leaves an unexplained increase in morbidity in NAFLD patients, which is caused by alternative conditions. Recent studies have found associations between NAFLD and specific extrahepatic cancers, including colon and breast cancer, as well as overall cancer risk.[59,202] The aim of the current chapter is to investigate the relationship between cancer incidence in NAFLD, and how this affects morbidity and mortality in NAFLD patients.

The relationship between NAFLD and cancer, as well as the synergy between NAFLD and other cancer risk modifiers is not fully understood. Hepatocellular carcinoma has long been associated with NAFLD and is widely recognised as one of the most severe endpoints of NAFLD.[201,203,204] The intracellular environment created by the presence of NAFLD has been found to contribute to HCC in a number of ways.[205] Damage to liver cells via oxidative stress, inflammation, and disruption of cytokines, adipokines and lymphokines are among the ways in which NAFLD contributes to the development of HCC.[206,207] There is evidence that NAFLD has effects on cancer incidence that extend beyond the liver.

Increased incidence of colorectal cancer has been found in patients with NAFLD in multiple studies; the first extrahepatic cancer found reliably associated with NAFLD.[208–213] A number of these studies found increased colorectal cancer risk in those with NASH compared to NAFLD.[210,214] Other sites linked to NAFLD include pancreas, oesophagus, stomach, breast, uterus, lung, ovary, and prostate.[59,215–218] A number of these studies have found even in non-

obese or non-overweight patients, that NAFLD is still associated with specific extrahepatic cancers.[59]

A small number of studies have found increased incidence of overall cancer risk in those with NAFLD. Kim et al. found markers of hepatic fibrosis to be associated with all cancer, and NAFLD itself to be associated with specific extrahepatic cancers; including breast and colon.[202] Allen et al. found increases in all cancer incidence in those with NAFLD.[59] The evidence for NAFLD being associated with all cancer risk is limited in comparison compared with the evidence for association between NAFLD and specific cancer sites. Further, large scale study is required to determine the true relationship between NAFLD and extrahepatic cancer.

To investigate the link between NAFLD and overall cancer incidence, it is important to disentangle NAFLD from other correlated cancer risk factors. Obesity, commonly defined as BMI equal or higher than 30kg/m^2, is a major cause of NAFLD, with 51.3% of NAFLD patients also being obese.[41,219] Obesity has also been linked with cancer incidence at 13 different sites in the body; a number of which have been reported to be associated with NAFLD also.[220] Wolin et al. estimate that excess weight or obesity accounts for 20% of all cancers.[221] Mechanistically, several factors associated with increased fat mass have been proposed to cause cancer.[222] For example, dysregulation of circulating hormones and cytokines including insulin, insulin–like growth factor signalling, adipokines, inflammation and sex hormones may disrupt normal cell cycle control and promote tumour formation.

Indeed, there is significant overlap of many of such pathological abnormalities between both overweightness and NAFLD.[223] The elements of shared pathophysiology of NAFLD and overweightness could potentially mean that the observed increases in cancer risk share a common aetiology. The increased risk of cancer incidence attributable with NAFLD and

obesity must be quantified to understand the role of each in previous associations observed. Allen et al. found that obese patients who did not have NAFLD were at only slightly higher risk of cancer incidence than those who were non-obese.[59] They also found that patients who were not obese, but had NAFLD were at increased risk compared to non-NAFLD controls. This led them to conclude that the majority of the observed cancer risk in obese patients is due to increased rates of NAFLD. It was found however that the effect of NAFLD compared to obesity differed depending on which cancer site was analysed. This may suggest a synergy between NAFLD and obesity for cancer risk.

Leading on from previous research in the area, the aim of this study was to analyse the effects of NAFLD on overall and specific cancer incidence. In addition to this, the study aimed to investigate the roles of BMI and NAFLD in increased risk cancer incidence. This study also investigated whether the relationship between NAFLD and cancer was causal or not.

3.3 Methods

3.3.1 Data

3.3.1.1 GoDARTS

This study aimed to analyse the incidence of all cancer longitudinally. The first cohort used was GoDARTS, a case-control type 2 diabetes study based in Tayside, Scotland. Key descriptive statistics and demographic attributes of this cohort are shown in Table 4.

Table 5 - Mean Characteristics of GoDARTS Patients Stratified by NAFLD Status at Time of Enrolment to GoDARTS

Characteristic	Non NAFLD	Number NAFLD	p
Number	6726	6969	
% Diabetic	24.65%(n = 1,658)	66.34%(n = 4,623)	< 0.0001
BMI	27.22kg/m^2(SD = 4.76)	30.90kg/m^2(SD = 6.07)	< 0.0001
Weight (Males/Females)	84.44kg(SD = 14.18)	92.89kg(SD = 17.04)	< 0.0001

	/70.20kg (SD = 14.38)	/79.64kg(SD = 18.30)	/< 0.0001
Female	41.48% (n = 2790)	55.46% (n = 3865)	< 0.0001
Smoker	55.14% (n = 3709)	57.67% (n = 4019)	$2.3x10^{-3}$
Age at Signup	61.88 years(SD = 13.72)	64.9 years(SD = 11.54)	< 0.0001
Follow-up Length	9.24years(SD = 2.50)	8.58 years(SD = 2.88)	< 0.0001

This cohort was used for discovery and comprised of electronic health records (EHRs) from 13,695 eligible individuals.[224] The mean age at sign up was 63.41 years and participants had a mean follow up of 8.95 years. 48.6% of patients were male. On patients' date of sign up they were phenotyped by biochemical and haematological investigations, anthropometric measurements and lifestyle questionnaires. This date was used as the beginning of the follow up period. 2,794 patients had cancer incidents during the follow-up period.

3.3.1.2 SHARE

Two further, independent cohorts were analysed for replication. The second data source was SHARE. This is a cohort in which individuals volunteer to allow their medical records to be used for scientific research, and is open to anyone in Scotland over the age of 16. The characteristics of this cohort are shown in table 5 below.

Table 6 - Mean Characteristics of SHARE Patients Stratified by NAFLD Status at Age 60 (Beginning of Follow-Up Period)

Characteristic	Non NAFLD	NAFLD	p
Number	19035	7856	
% Diabetic	1.65% (n = 314)	8.17% (n = 6418)	< 0.0001
Female	52.45% (n = 9984)	65.10% (n = 5114)	< 0.0001
Follow-up Length	13.92 years(SD = 7.45)	6.91 years(SD = 4.31)	< 0.0001

This comprised 62,438 patients with EHRs available once patients with exclusions for alternate causes of raised ALT livers were removed.[188] This cohort was used for replication of findings in GoDARTS. The mean age in SHARE was 57.0 years, and 61.6% were female.

3.3.1.3 Tayside and Fife

Replication of results was also undertaken in Tayside and Fife diabetics (T&F). This cohort comprises all patients in the Tayside and Fife NHS region who had a diagnosis of T2DM at some point in their lives. Many of the patients received a diagnosis of T2DM during the follow up period, therefore the T2DM rate is not 100% at baseline. The characteristics of this cohort are shown in table 6.

Table 7 - Mean Characteristics of Tayside and Fife Diabetics Patients Stratified by NAFLD Status at Age 60 (Beginning of Follow-Up Period)

Characteristic	Non NAFLD	NAFLD	p
Number	5,102	6039	
% Diabetic	40.53% (n = 2068)	60.37% (n = 3646)	< 0.0001
BMI	30.55kg/m^2 (SD = 6.12)	33.15kg/m^2 (SD = 6.78	<0.0001
Female	45.84% (n = 2339)	53.27% (n = 3217)	< 0.0001
Smoker	58.57% (n = 2988)	65.01% (n = 3926)	<0.0001
Follow-up Length	11.08 years (SD = 6.07)	6.21 years (SD = 4.03)	< 0.0001

Like the two previous cohorts, medical records from the NHS are available for these patients. The cohort 16,312 patients eligible after exclusions for other hepatic insults were made. The mean age of these patients was 65.0 years, and 48.1% were female.

The results from T&F were not meta analysed with GoDARTS and SHARE, as this is a primarily diabetic cohort, therefore does not capture those who do not go on to get diabetes.

The ascertainment bias in this cohort that only contains individuals who did eventually get diabetes is likely to have resulted in the lower point estimate for NAFLD in cancer risk that we have observed. To ensure there was no overlap in participants between cohorts, patients in SHARE who are also in GoDARTS were excluded from SHARE, and participants in GoDARTS or SHARE were excluded from analysis in T&F, meaning each cohort is completely independent.

To allow comparison with the GoDARTS cohort, a baseline point had to be chosen from which to begin the follow up period in which to analyse cancer incidence in SHARE and T&F. The age of 60 was chosen as it is close to the mean baseline age of GoDARTS, and importantly close to the mean age of NAFLD diagnosis in GoDARTS (60.8 years) and in the literature.[1] This allowed a more robust replication of findings in GoDARTS in the two replication cohorts, ensuring age wasn't a source of heterogeneity in analysis. These criteria left 26,891 patients in SHARE and 11,141 patients in T&F suitable for analysis with a median follow up time of 11.0 years and 8.0 years respectively. The EMRs available for patients in all cohorts are from the NHS Tayside and Fife authorities.

3.3.2 Outcomes

All outcomes were defined using NHS medical record data, made available for participants in each of the three cohorts. As such, all data was recorded in the same format; all disease was recorded in ICD10 codes and biochemical measures in the same, relevant units.[225] (E.g. Units per litre for ALT measurements)

3.3.2.1 NAFLD Phenotype

NAFLD cases and controls were defined using the liver function test alanine transaminase (ALT), a commonly used marker of liver damage, and a useful surrogate for NAFLD.[131] A full description of this phenotype is given in a previous chapter. This was chosen as it is

commonly measured, and a large portion of the population have multiple measurements. Elevated ALT levels were considered to be over 30U/L for men, and over 19U/L for women (Normal ALT reference ranges: Males - 5-30U/L; Females – 5-19U/L.). These upper limits are those suggested by Prati et al. as the maximum normal values of ALT in healthy adult men and women.[180] Raised ALT levels correlate with NAFLD and are an appropriate surrogate marker for the disease, provided other causes of liver disease are ruled out.[131,226] There is substantial evidence that raised ALT levels in the absence of any apparent liver insult are extremely likely to be caused by NAFLD.[227] All samples from GoDARTS, SHARE and T&F were analysed in the same laboratory.

NAFLD cases were defined as any patient who had experienced at least 2 raised ALT measurements, at least 3 months apart. This time scale was chosen as 3 months is a commonly used definition of chronic and most cases of acute hepatitis, such as drug induced, will have resolved.[228] This also increases the specificity of the definition.

3.3.3 Exclusions

Patients were excluded from analysis if they had features of other chronic liver disease recorded in their medical records. These included: any positive serological tests for anti-smooth muscle antibody, antinuclear antibodies or anti-mitochondrial antibodies, any positive serology for hepatitis B surface antigen or hepatitis C antibody, or mention of cause of liver disease in medical records. In GoDARTS, 1,157 patients had both immunological and virological screens at some point which were negative, therefore they were included in analysis. Patients with alcohol dependence or any documentation of alcoholic liver disease in their EHRs were excluded using ICD codes: “K70” and “F10”. In addition, patients who self-reported drinking more than 20g (2.5 units) a day for women and more than 30g (3.75 units) a day for men were excluded. Allen et al. concluded that alcohol was not likely to explain the increase in cancer incidence seen with NAFLD in their study.[59]

3.3.4 Validation of Phenotype

To validate this NAFLD phenotype, sensitivity and specificity analyses were conducted in GoDARTS comparing this to cases of NAFLD confirmed in EHRs with the "K76.0" ICD10 code. Full validation of this phenotype is included in the previous chapter. The sensitivity of this definition was 97.4%, and the specificity was 32.0%. These analyses were also conducted in SHARE using the same method, with a sensitivity of 75.3% and specificity of 54.2%, and in T&F with sensitivity of 94.6% and specificity of 38.3%. The SHARE cohort has lower sensitivity compared to the other two cohorts, likely due to the lower average age of the cohort and the lower prevalence of diabetes resulting in lower healthcare interaction, morbidity and mortality. Also, due to the relatively low numbers of confirmed NAFLD in EHRs, small differences in numbers can have large effects on sensitivity and specificity.

To further validate this phenotype positive control tests were run against chronic kidney disease (CKD) in GoDARTS, as it has been shown to associate with NAFLD.[189] During the follow up, 1,131 patients had incidence of CKD. NAFLD was found to associate with incidence of CKD in a cox proportional hazards model adjusted for sex, T2DM, age, and BMI. (HR = 1.32(1.25 – 1.39), $p = 4.8x10^{-5}$)

A positive control test with the well-known NAFLD risk variant *PNPLA3* rs738409 was conducted. [190] In GoDARTS, 8,399 eligible participants had been genotyped for this variant. In a logistic regression with an additive model, adjusted for age and sex, *PNPLA3* rs738409 was associated with increased NAFLD at the beginning of the study. (OR = 1.32(1.12-1.36), $p = 1.09x10^{-8}$)

3.3.5 NAFLD Related Phenotype Definitions

As well as our ALT based NAFLD definition, some patients had NAFLD confirmed in hospital admissions data with the ICD10 code "K76.0". This is referred to as "NAFLD

hospitalisation" in subsequent sections. In GoDARTS, 0.36% of participants had this code reported in their medical records at any point.

Non-Alcoholic Steatohepatitis (NASH) was phenotyped by searching admissions, deaths and biopsy files for cases of NASH, defined using the ICD10 codes for NASH, fibrosis and cirrhosis. This may have been a main cause of hospitalisation or concomitant morbidity.

Another method of detecting NAFLD non-invasively is the Fatty Liver Index (FLI).[148] This uses BMI, waist circumference, triglycerides and gamma-glutamyl transferase (GGT) to define NAFLD, and has been validated in a number of cohorts as an accurate surrogate of NAFLD. 4,164 patients in GoDARTS had the required data available for this measure. In GoDARTS, FLI correlated significantly with NAFLD as diagnosed by ALT levels. (Pearson correlation coefficient = 0.33 (0.31-0.36), $p < 0.0001$)

The FIB-4 scoring system was also used in the GoDARTS study.[153] A FIB-4 score of greater than 3.25 has been shown to predict advanced hepatic fibrosis, therefore this score was used as the cut off. This was calculated using the highest recorded AST and ALT measurements and platelet count before the beginning of the GoDARTS for each individual to calculate the highest FIB-4 score they had experienced.

3.3.6 Mendelian Randomisation

Mendelian randomisation methods were used to assess whether the relationship between NAFLD and cancer incidence was causative.[229] The missense variant *PNPLA3* rs734809, which is strongly associated with the development and progression of fatty liver disease, was chosen as it has been shown in a large number of studies to associate with NAFLD, and has been used in previous Mendelian randomisation studies on NAFLD. The ratio method was used to conduct this analysis.[190] In GoDARTS, 7,715 patients had been genotyped for this variant, and 343 of these were homozygous carriers. (Minor Allele Frequency (MAF) =

20.6%) In SHARE, 1,755 patients had been genotyped for this variant, with 50 being homozygous carriers. (MAF = 23.0%)

3.3.7 Overweight and Obesity Definitions

In this study, overweight is defined as a BMI greater than 25kg/m^2 and less than 30kg/m^2. Obesity is defined as a BMI equal or over 30kg/m^2.[230]

3.3.8 Cancer Phenotype

Cancer incident data was obtained from the Scottish cancer register, part of the Scottish Morbidity Record.[231] This contains all diagnoses of cancer made in Scotland in NHS care, in ICD10 code format. This data was available for patients in GoDARTS, SHARE and T&F. Cases were cross checked with recorded cases in hospital admissions and death records files. The cancer records were identified by the relevant ICD10 codes for malignant neoplasms or neoplasms of unknown behaviour. These were any code including "C", "D0", "D37", "D38", "D39", or "D4". Obesity related cancer incidents were phenotyped similarly, but specifically for the 13 reported obesity related cancer sites. [220]

Cancer deaths were phenotyped based on death certificate files in EHRs. These list a main cause of death and contributing causes of death for each patient who has died. These were also cross checked with the Scottish cancer register file.

3.3.9 Statistical Methods

All data analysis was carried out in the statistical package R. The effects of NAFLD and other independent variables on cancer incidence were analysed using a Cox proportional hazards model (CPH). Patients were censored at the point at which they had a cancer incident recorded, death, or September 2016 when the follow-up period ended. Patients with missing data were excluded from analysis.

To assess whether NAFLD affected cancer death risk in the presence of non-cancer death as a competing risk regressions (CRR) using Fine and Gray's method were run. Logistic regression (LR) models were used to evaluate the effect of NAFLD on death cause.

In the GoDARTS cohort models were adjusted for sex, age, BMI, T2DM, and smoking status. In GoDARTs, models with BMI replaced by weight or waist measurement were also run, as these are slightly different measure of obesity and may have provided further insight into the associations. Hypertension, activity level, alcohol consumption and deprivation level were not included in the models as they did not have a significant effect on cancer incidence in the adjusted model. In the SHARE cohort, models were adjusted for sex and T2DM. Smoking and BMI data were not widely available for individuals in the SHARE cohort, therefore this was not controlled for in most analyses.

3.4 Results

3.4.1 NAFLD and Cancer *Incidence*

In the GoDARTS cohort, NAFLD was associated with increased cancer incidence. During the follow up period, 18.5% of controls compared to 22.2% of patients with NAFLD

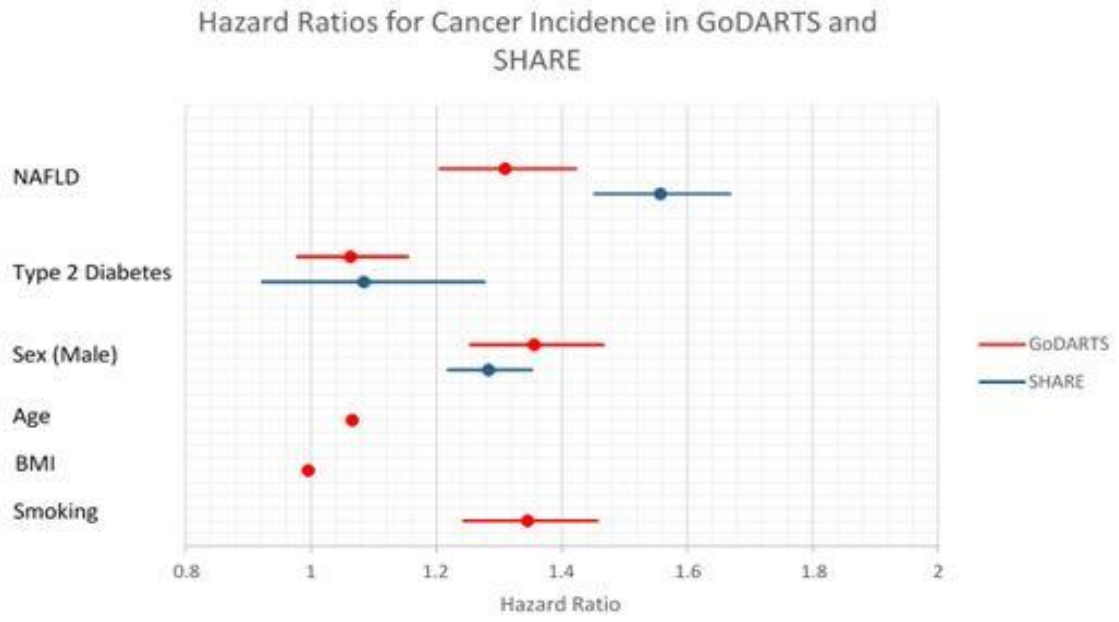

Figure 3-1Forest Plot of Effects of Variables on Cancer Incidence in GoDARTS and SHARE

developed cancer. In controls, 1244 patients had cancer incidents and 1550 patients had incidents in NAFLD cases. Patients who had NAFLD at enrolment to GoDARTS had increased cancer incidence independent of sex, age, BMI, smoking status, and diabetes status (HR = 1.31(1.27 – 1.35), $p = 1.8x10^{-10}$). These results are shown in Figure 3-1 below.

Using the same covariates, the Fatty Liver Index was associated with increased cancer incidence (HR = 1.004(1.00-1.008), $p = 5.0x10^{-2}$) and FIB-4 score over 3.25 was associated with increased cancer risk. (HR= 1.31, 95% CI =1.29 – 1.53, $p = 3.2x10^{-3}$).

When NAFLD was not taken into account, BMI was associated with increased cancer incidence (HR = 1.09(1.01 – 1.18), $p = 3.1x10^{-2}$). This association was completely abrogated when adjusted for the presence of NAFLD. Similar results were found for other markers of adiposity, weight and waist measurements.

When analysis was limited to obesity related cancers, BMI was associated with increased cancer incidence. (HR = 1.01(1.00 – 1.03), $p = 3.3x10^{-2}$) Similarly to the analysis of all cancer incidence, BMI was not associated with cancer incidence when NAFLD was added as a covariate.

Similar results were found in the SHARE cohort. Out of 26,891 patients analysed, 5,728 had cancer incidents in the follow up period. NAFLD was associated with increased cancer incidence (HR = 1.56, (1.45- 1.67), $p < 2x10^{-16}$). NAFLD hospitalisation prior to baseline was associated with increased cancer risk, with a hazard ratio of 2.54. (95% CI = 1.14 – 5.65, $p = 2.3x10^{-2}$). NASH was also associated with increased cancer incidence. (HR = 4.18(1.74-10.0), $p = 1.4x10^{-3}$) Among the patients in SHARE, 1,912 had BMI data available. In these patients, when NAFLD was accounted for, BMI was not significantly associated with overall cancer incidence, or with obesity related cancer incidence.

Similar results were found in the population based diabetes cohort from Tayside & Fife. Out of the 11,141 patients analysed, 1,819 had cancer incidents in the follow up period after the age of 60. NAFLD was associated with cancer incidence in the follow up period. (HR = 1.16(1.04-1.29), p =$5.9x10^{-3}$)

As well as increasing all primary cancer incidence, NAFLD was associated with increased incidence of specific cancers in GoDARTS and SHARE, shown in Figure 3-2 below.

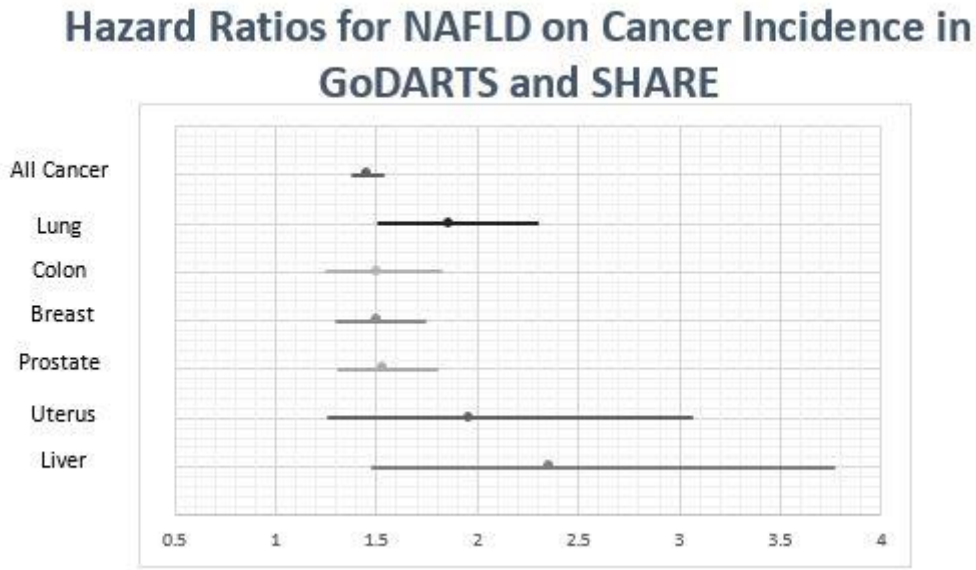

Figure 3-2 Hazard Ratios for Effect of NAFLD on Specific Cancer Sites in GoDARTS and SHARE Meta-Analysis

Due to lower numbers of cases, the confidence intervals for these are wider than for all primary cancers combined. Breast and uterine cancer analyses were limited to females, with prostate cancer analyses limited to males. T&F was not meta analysed in this analysis due to the primarily diabetic composition of the cohort, which did not capture those over 60 who did not go on to get T2DM. The ascertainment bias in this cohort that only contains individuals who did eventually get diabetes is likely to have resulted in the lower point estimate for NAFLD in cancer risk that we have observed.

In the T&F cohort of the 11,141 patients analysed, 1,819 had cancer incidents in the follow up period after the age of 60. NAFLD was associated with cancer incidence in the follow up period in an adjusted cox proportional hazards model. (HR = 1.16(1.04-1.29), p =$5.9x10^{-3}$) NAFLD hospitalisations were significantly associated with cancer incidence in the same model. (2.04(1.12-3.71), p = $1.9x10^{-2}$) BMI was not associated with cancer incidence when NAFLD was included in the analysis. When analysis was limited to obesity related cancers, BMI did not show any significant association with cancer incidence when NAFLD was adjusted for.

3.4.2 NAFLD and Cancer Death

The relationship between NAFLD and cancer death was analysed in GoDARTS. In a CPH model adjusted for age, sex, diabetes, BMI and smoking, it was found that NAFLD is associated with increased risk of cancer death. (HR = 1.40(1.21-1.61), p = $8.8x10^{-4}$) FLI was associated with increased cancer death risk in the same CPH model. . (HR = 1.009(1.002 - 1.015), p =$9.8x10^{-3}$)

NAFLD was associated with increased risk of non-cancer death in the same model. (HR = 1.23(1.12-1.35), p < 0.0001) To estimate the effects of NAFLD specifically on cancer death more accurately, competing risks analyses were run.

A CRR using Fine and Grays's method was run to analyse the association between NAFLD and cancer death with non-cancer related death as a competing risk. In a model with sex, T2DM, smoking, obesity and age, NAFLD increased risk of cancer with a subdistribution hazards ratio (SHR) of 1.28 (95% CI =1.11 - 1.47, p = $8.8x10^{-4}$).

In SHARE, a CRR adjusted for sex and T2DM with non-cancer death as a competing risk was run. Patients with NAFLD had a significantly higher risk of cancer death. (SHR = 3.12 (2.38– 4.10), p <$2x10^{-16}$)

In T&F, in an adjusted competing risks regression with non-cancer death as a competing risk, NAFLD was associated with increased cancer death. (SHR = 1.40(1.20-1.63), $p = 9.6x10^{-11}$)

In those patients who died during the follow up period of GoDARTS, NAFLD was associated with increased chance of cancer being the main cause of death in a logistic regression adjusted for age, sex, T2DM, smoking and BMI (OR = 1.33 (1.10 – 1.62), $p = 3.6x10^{-3}$). This was also found in SHARE in a logistic regression adjusted for sex and T2DM. (OR = 1.54(1.17 – 2.03), $p = 2.0x10^{-3}$)

The same result was found in T&F, with patients with NAFLD more likely to die with cancer as the main cause of death. (OR = 1.44(1.32 – 1.58), $p = 1.3x10^{-3}$)

3.4.3 Cancer as a Driver of Shorter Lifespans in NAFLD Patients

Further analysis showed that this association between NAFLD and cancer death is one of the major drivers of the shorter life expectancies of NAFLD patients. This is shown in the table 7 below.

Table 8- Proportion of All Deaths Due to Cancer Stratified by NAFLD and Type 2 Diabetes Status in GoDARTS

Group	Cancer as Main Cause of Death Rate	Cancer as Contributing Cause of Death Rate	Total Number in Group
No NAFLD or T2DM	31.33%	35.34%	382
T2DM	24.28%	30.29%	449
NAFLD	41.25%	45.26%	559
Both NAFLD and T2DM	27.80%	31.73%	1853

Table 7 shows the proportion of all deaths which are attributable to cancer. For example, in patients with NAFLD and not diabetes, 41.25% of deaths had cancer a the main cause, and 45.26% of all deaths in this group had cancer as a main or contributing cause.

In GoDARTS, when stratified by cancer death and non-cancer death, NAFLD had no effect on age of death in the non-cancer group. NAFLD was associated with lower death age in those patients who died with cancer as a main cause. ($p = 6.1x10^{-5}$, $\beta = -2.91$, 95% CI= (-2.18, -3.63), adjusted $R^2 = 0.05$) NAFLD did not have an effect on age of death in those who never had a cancer diagnosis, but associated with lower age of death in those who had a cancer diagnosis at some point. ($\beta = -2.07$, 95% CI= (-1.54, -2.60), adjusted $R^2 = 0.07$, $p = 1.8x10^{-4}$)

Mean age of death is presented in table 8 below. This is stratified by NAFLD, T2DM and cancer death.

Table 9 - Mean Death age versus Cancer Death, NAFLD and Type 2 Diabetes. (● *indicates condition is present*)

Cancer as Main Cause of Death	NAFLD	T2DM	Mean Death Age	N
			82.7	263
●			79.2	119
	●		83.0	328
●	●		75.9	231
		●	80.8	340
●		●	79.9	109
	●	●	79.4	1340
●	●	●	76.6	513

Similar results were found in SHARE, although these were less comparable due to the younger age of the SHARE cohort therefore lower numbers. NAFLD was associated with lower age of death in those who had a cancer diagnosis at some point in their life ($\beta = -2.04$, 95% CI= (-0.25, -3.84), adjusted $R^2 = 0.07$, $p = 0.026$), but not in those who never had a cancer diagnosis. This is shown in table 9 below.

Table 10 - Mean Death age versus Cancer Death, NAFLD and Type 5 Diabetes. (● indicates condition is present)

Cancer as Main Cause of Death	NAFLD	T2DM	Mean Death Age	N
			75.5	116
●			73.2	103
	●		76.4	270
●	●		69.8	332
		●	72.4	11
●		●	70.7	5
	●	●	73.6	89
●	●	●	71.6	54

3.4.4 PNPLA3 and Cancer Incidence

The effects of *PNPLA3* on cancer incidence during the follow up period in GoDARTS and SHARE were evaluated. Homozygous carriers of *PNPLA3* rs738409 had increased risk of cancer incidence (HR = 1.27 (1.02-1.58), $p = 3.1 \times 10^{-2}$). These results were meta-analysed with results from SHARE, shown in Figure 3-3.

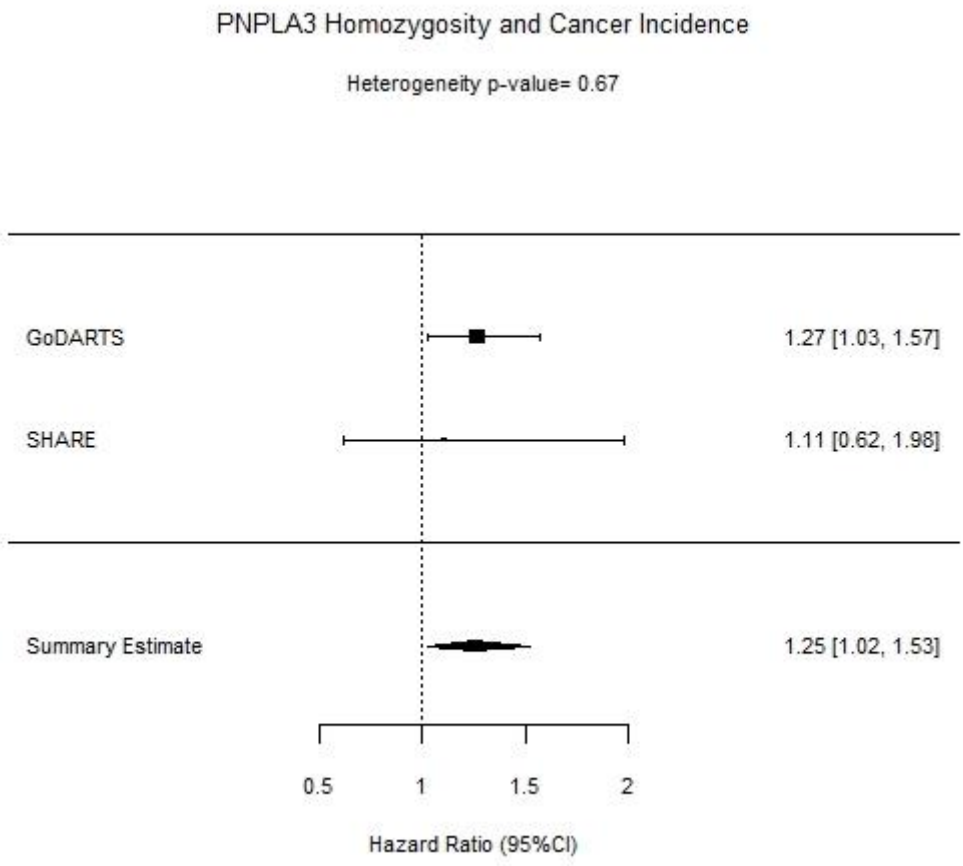

Figure 3-3 Forest Plot of Effects of PNPLA3 rs738409 on Cancer Incidence in GoDARTS, SHARE and Meta-Analysis

This association was also observed in GoDARTS when patients with liver cancer were excluded from analysis, as *PNPLA3* rs738409 has been shown to increase liver cancer risk.[232] (HR = 1.26(1.01-1.58), $p = 3.8x10^{-2}$) Similar results were found in an adjusted CRR with death as a competing risk. (SHR =1.24(1.00-1.54), $p = 4.9x10^{-2}$)

Mendelian randomisation analysis was conducted to estimate the effect of NAFLD on cancer incidence. Using the ratio method in a meta analysis of GoDARTS and SHARE, NAFLD was found to be significantly associated with cancer incidence, with a β estimate of 1.33(95% CI = 0.18 - 2.49, p = 0.023)

3.5 Discussion

In this study, it is shown that a significant increase in cancer incidence exists in patients with NAFLD. Cancer incidence and death was higher in those who had NAFLD in GoDARTS, SHARE, and T&F using the raised ALT definition as a surrogate of NAFLD. This demonstrates the generalisability of this result. This is the first truly large scale observational study to show these associations, as well as the first to show the effect of BMI on cancer incidence is driven to null when NAFLD is accounted for. In SHARE and T&F, NAFLD admissions were associated with increased cancer incidence. NASH also increased cancer incidence, with a larger effect size than NAFLD. Other non-invasive biomarkers including Fatty Liver Index and FIB-4 score prior to enrolment to the GoDARTS study were also found to increase risk of cancer during the follow-up period. These results support findings from other published studies that link NAFLD to cancer of all types. [59,202] It also suggests that the more pro-inflammatory form of NAFLD, NASH, may have more of an effect and this may give clues to the biological mechanism(s).

There is emerging evidence that the association between NAFLD and cancer extends beyond the liver to other parts of the body. Kim et al. found in a cohort follow-up study that, in addition to an increased risk of liver cancer, NAFLD dramatically increased rates of extrahepatic cancers, including breast and colon in those who were diagnosed with NAFLD prior to the 10 year follow-up period.[202] Allen et al similarly showed that NAFLD was associated with increased extrahepatic cancer risk, in sites such as the colon, lung and prostate.[59] In the current study, an increase in cancer incidence was found in many of these specific sites, including breast, colon, liver, lung and prostate. Collectively, these data, including the results that we describe, supports the notion that NAFLD increases incident cancer risk.

NAFLD is associated with increased risk of cancer death in the follow up period in both GoDARTS, SHARE and T&F. This data correlates with our earlier findings that NAFLD is associated with increased cancer incidence, as increased incidence is naturally linked to increased mortality. Analysis of causes of death as reported by ICD10 codes in medical records showed that the deaths of patients with NAFLD were more likely to be due to cancer. Cancer was a key factor in the shorter lifespans of patients with NAFLD, as there was no significant effect of NAFLD on age of death when patients with a cancer diagnosis were excluded. This increase in cancer incidence and death accounts for a large proportion of the increase in morbidity and mortality shown in NAFLD patients in the previous chapter. Similar results were found in a recent study in a large Swedish cohort with biopsy confirmed NAFLD.[60] In this study, Simon et al. found that excess death in NAFLD patients was primarily driven by extra-hepatic cancers and cirrhosis, while other causes such as cardiovascular disease and HCC had only a small effect. These findings agree with those of the current study, further implicating NAFLD in the development of extrahepatic cancer.

We showed that homozygous carriers of the *PNPLA3* NAFLD risk variant, rs738409, had an increased risk of cancer incidence. In a Mendelian randomisation analysis, we showed *PNPLA3* rs738409 increased NAFLD incidence, NAFLD increased cancer incidence, and *PNPLA3* rs738409 increased cancer incidence. This novel finding is supporting evidence that NAFLD is causally associated with increased cancer incidence.

Substantial evidence links cancer to hyperinsulinemia. For example, hyperinsulinemia has been found to be a risk factor for colon cancer.[233] It is also a risk factor for cancer death, independent of obesity.[234] Patients with NAFLD are more likely to have hyperinsulinemia, and this is associated with reduced insulin clearance.[235] This insulin excess may underlie, at least in part, the mechanistic basis by which NAFLD increases cancer incidence, as insulin/Igf-1 may promote tumour formation through mitogenic pathways downstream of

their receptors.[236] No association was seen between T2DM and cancer when NAFLD was accounted for however, which suggests factors other than hyperinsulinemia are causing the increased cancer.

Damage to liver cells via oxidative stress, inflammation, and disruption of cytokines, adipokines and lymphokines may contribute to the pathogenesis of cancer in those with NAFLD.[205] NAFLD is a pro-inflammatory state, which may create an environment favourable to the development of cancer.[237] Cancer has been previously linked to chronic inflammation, via increases in mutations, reduced apoptosis and other environmental changes.[238] Inflammatory mediators suck as arachidonic acid, cytokines, chemokines, and free radicals are increased in NAFLD, and may contribute to cancer risk.[238] Adipokines and cytokines for example are found in higher concentrations in the serum of NAFLD patients, which may be factor in extrahepatic cancer development.[239] Pro-inflammatory cytokines TNF-α (Tumour Necrosis Factor alpha) and IL-6 (Interleukin-6) are key factors in NAFLD inflammation, and may encourage development of neoplasia.[238] Due to the diverse nature of cancer at different sites, it is likely that a number of different factors have effects. The larger effect sizes of NASH and FIB-4 on cancer incidence observed in this study are consistent with the notion of inflammation driving a proportion of cancer risk, as compared to simple steatosis.

In a model adjusted for age and sex, BMI was found to be associated with increased cancer incidence. Many studies have shown increased cancer risk with increasing BMI, therefore this finding is consistent with previous literature. We found that BMI was not associated with overall cancer incidence when NAFLD was taken into account, and the same was found for waist and weight measurements. We also found that individuals who were obese but did not have NAFLD were not at increased risk of cancer incidence compared to those of a healthy weight. This finding supports those of Allen et al.[59] When analyses were limited to so-called

obesity related cancers, similar results were found, as BMI was associated with cancer incidence, but not when NAFLD was adjusted for. This was found in all three cohorts analysed. The lack of independent association between BMI and cancer incidence in our study may suggest that NAFLD is a major component in the increased risk of cancer observed in overweight and obese patients.

These findings have a number of implications for different stakeholders. For patients, especially those who are overweight or have T2DM, this has the most impact. The association between ALT and future cancer incidence may be useful as a screening tool to identify those with increased cancer risk. Methods of screening for cancer can be invasive and uptake is often low, therefore a blood based biomarker which is routinely measured in many individuals could be an effective adjunct to current screening methods such as mammograms and colorectal screening.[240–242]

This finding increases the importance of creating a pharmacological intervention for NAFLD. The findings of this study combined with other studies linking NAFLD to extrahepatic cancer may also aid cancer research in pinpointing the pathways and pathologies which link excess body weight to cancer incidence.

3.5.1.1 Limitations

NAFLD Phenotype

The NAFLD phenotype may be a limitation of this study, and this is discussed in detail in the previous chapter. There is substantial evidence linking ALT levels to NAFLD, but also evidence that NAFLD can exist in patients with normal ALT levels. Furthermore, non-sensitive NAFLD phenotype would drive the association towards null, and therefore we cannot exclude the fact that the true association may be stronger than that we have observed. While we acknowledge that ALT levels may have a limited sensitivity for defining mild

NAFLD, we have shown that our ALT based definition is highly sensitive for more advanced cases, such as those with the Fatty Liver Index measured and those hospitalised with steatosis. In GoDARTS, SHARE and T&F, we estimated sensitivity to be 97.4%, 75.3% and 94.6% respectively for such advanced cases.

Genetic evidence for the suitability of our phenotype ascertainment is demonstrated by the observation that the major NAFLD susceptibility variant in *PNPLA3,* rs738409, was associated with our NAFLD phenotype with a very similar magnitude to that previously reported. [191] The high sensitivity of the phenotype and similar effects of other NAFLD related phenotypes on cancer incidence, plus previous literature linking NAFLD to cancer support the validity of the ALT based NAFLD phenotype. [59,202]

Whilst we show that our NAFLD phenotype is accurate, even if part of the aetiology of the raised ALT levels is alcohol or another cause, this is still an important and interesting result. The observation that when ALT levels are taken into account, BMI no longer associates with cancer incidence changes current understanding of the link between cancer and obesity. The association between raised ALT levels and future cancer incidence, even agnostic of the aetiology, is a valuable finding which may be used for cancer risk screening and prediction.

We found NASH to be associated with increased cancer incidence, and suggest its associated hepatic inflammation may contribute to cancer risk. The majority of patients with NASH however also have a diagnosis of fibrosis, which could mean the effect is fibrosis rather than inflammation driven.

Covariate Data Missingness

The missingness of BMI and smoking data for patients in SHARE is a possible limitation of the current study. In the analysis of cancer incidence in GoDARTS, T&F and the sub-group of SHARE patients with BMI data available, the inclusion of BMI as a covariate did not

modify the association between NAFLD and cancer. In GoDARTS also, NAFLD was not associated with rates of smoking when age and sex were adjusted for. Due to this, the analysis of NAFLD and cancer without BMI and smoking as covariates is still valid, and comparable with the analyses undertaken in GoDARTS. Allen et al, used similar methodology, as they did not correct for smoking and found that BMI played a relatively small part in cancer risk compared to NAFLD.[59] The self-reported nature of alcohol intake in GoDARTS, and missingness of this data in SHARE and T&F, as well as the ubiquitous nature of alcohol consumption at the sub clinical level, does not allow us to exclude the possibility that general alcohol consumption may play a role in the relationship between NAFLD and cancer, however this is likely to be a limitation of the concept of NAFLD in general.

BMI as a marker of Obesity

BMI is an accurate and useful marker of obesity, although is not perfectly correlated with abnormal body fatness as factors such as muscle mass can impact the result.[243] To assess whether this was a factor in the lack of association between BMI and cancer incidence, other measures of body fatness including waist measurement and weight were analysed. These also did not associate with cancer incidence when NAFLD was considered. Though other markers of excess body fatness exist, such as waist to hip measurements can be useful techniques, the fact BMI, weight nor waist measurement associated with cancer risk when NAFLD was considered suggests that this did not significantly impact findings.

3.6 Conclusion

In the current study we have shown that NAFLD is associated with increased risk of cancer incidence. There is also an association between NAFLD and cancer death, and cancer is a key factor in the shorter life expectancies associated with NAFLD patients. Furthermore, we are

first to show the association between BMI and cancer is driven to null when NAFLD is included in the model. This is further replicated in two additional, large cohorts, demonstrating the robust nature of this relationship. Given the large numbers of participants, these findings are likely generalisable to the general population. A key, novel finding of the study was that the missense variant *PNPLA3* rs738409 is associated with increased cancer incidence. These findings suggests that the effect of NAFLD on cancer incidence may be causative, and that a major component of the association between body weight and cancer may be driven by NAFLD.

4 GWAS of NAFLD in Scottish and Indian Populations

4.1 Abstract

NAFLD is a common cause of liver disease and affects roughly a quarter of adults globally. There is a significant genetic component of NAFLD risk. The aim of this study was to find genetic determinants of NAFLD, and explore the effects of NAFLD variants in two Scottish and Indian study populations.

In this cross-sectional cohort study, genome wide analysis studies (GWAS) were run to find genetic determinants of NAFLD. The GoDARTS and SHARE cohorts from Scotland, and the DMDSC cohort from India were the sources of data. NAFLD was defined by the presence of elevated ALT levels. Fatty Liver Index (FLI) and Fatty Liver Grade defined by ultrasound (FLG) were also available in the GoDARTS and DMDSC cohorts respectively.

In both DMDSC and GoDARTs, *PNPLA3* rs738409 was associated with increased NAFLD risk, with an odds ratio (OR) of 1.34 ($p < 1x10^{-15}$). In GoDARTS, variants in the *CHUK*/*ERLIN1* locus were genome wide significant for NAFLD (OR = 0.89($p = 2.1x10^{-8}$). In GoDARTS, variants in *FAM19A4, EOGT, DNAH11,* and TCF7L2 were genome wide significantly associated with FLI.

We found that genetic variation had a significant contribution to NAFLD risk in both cohorts. *PNPLA3* rs738409 increased risk of NAFLD in both GoDARTS and DMDSC, concurring with previous research that this locus is a key component of genetic NAFLD predisposition globally.

4.2 Introduction

Non-Alcoholic Fatty Liver Disease is a common cause of liver disease and is rising in prevalence globally.[244] Research has revealed a number of genetic modifiers of NAFLD risk.[190,245] Study of the genetics of NAFLD are a key component in the investigation to

understand the disease, and may provide insights that aid drug discovery and the application of personalised medicine.[246,247]

Genetic studies of NAFLD have been undertaken extensively in European cohorts. The first GWAS of NAFLD by Romeo et al. found that the SNP *PNPLA3* rs738409 was associated with increased NAFLD risk, and this has been replicated a number of times.[190,248,249] Several other GWAS since have revealed many genetic loci which influence NAFLD, and many have found variants which have effects on steatosis, NASH, fibrosis and cirrhosis individually.[245,250–252]

A number of studies have investigated the genetics of NAFLD in Indians, but a search of GWAS catalogue found no results for GWAS studies in Indian subjects.[253] A number of candidate gene studies in Indian populations have found NAFLD related genetic variants.[254,255] Due to the large and diverse nature of the population of India, many studies have analysed differences in genetic modifiers between different groups and regions within the country.[256] Chatterjee et al. investigated the frequency of 34 known NAFLD risk variants, finding that they were overall more common in caste populations than tribal populations.[99]

Diversity of phenotypes and genetic predisposition to disease has been found between Indian and European populations previously. For example, risk of type two diabetes (T2DM) is known to be heterogeneous between Europe and India, and recent findings have suggested this may be due to genetics.[257,258] Studies have shown differences in a number of clinical parameters including obesity and heart disease risk.[259] Investigations into whether NAFLD is different between white Europeans and Asian Indians have been conducted, and suggested that increased NAFLD risk accompanies increased insulin resistance in Indian men.[260,261]

The aim of the current study was to identify genetic influences of NAFLD and NAFLD related phenotypes in Scottish and South Indian populations using data from electronic health

records. (EHRs) Further to this, the study aimed to compare the genetic modifiers in each cohort and evaluate differences in frequency between risk variants in each population.

4.3 Methods

4.3.1 Data

The analysis for this study was conducted in the GoDARTS and DMDSC cohorts, which are described fully earlier in the current thesis.

The first GWAS was conducted in Scottish data from the GoDARTS study.[187] This is a T2DM case-control cohort with electronic health record data available for the 18,306 participants. After exclusions were made for other causes of liver disease, there were 7,629 individuals eligible for analysis with sufficient clinical and genotype data available.

Further GWAS analyses were conducted in the Dr Mohan's Diabetes Specialities Centre (DMDSC) cohort.[195] This cohort comprised 75,952 individuals who were patients of the DMDSC and had T2DM. These individuals were predominantly South Asian. From this cohort, there were 3,154 individuals who had been genotyped and had sufficient data available for analysis.

4.3.2 Data Quality Control

A number sequential steps to ensure the validity and accuracy of the genetic data was undertaken. These were the same for all cohorts and genotyping platforms.[262] All steps were completed using Plink 1.9 software.[263]

Firstly, individuals with missingness of genetic data greater than 0.02 were excluded. SNPs with missingness greater than 0.02 were also excluded. SNPs with minor allele frequency (MAF) lower than 0.01 were excluded. SNPs which were not in Hardy-Weinberg equilibrium were excluded.[264] A p value threshold of $1x10^{-5}$ was used for this. Individual participants were excluded if they had a heterozygosity rate greater than three standard deviations from

the mean.[265] The SNPs were pruned for linkage disequilibrium (LD) using a sliding window technique. This window was 50 SNPs wide, moved 5 SNPs each step and variants with correlation greater than 0.2 in this window were excluded.[266] The cohorts were checked for relatedness between individuals and, those with pihat greater than 0.2 were excluded.[267]

Ethnic outliers were excluded from each cohort. Using multidimensional scaling of the genetic data, 20 principle components were calculated.[262] This was conducted for each cohort, plus the 1000 Genome cohort (1000G).[268] Each cohort was plotted with respect to the first two principle components against the 1000G cohort, which was labelled with ethnicity. Ethnic outliers were then removed from each cohort.

4.3.3 Measures

An in depth analysis and description of the phenotypes used in the current analysis is included in a previous chapter of the current thesis. A brief outline of each is included below.

In GoDARTS, two phenotypes were analysed. These are described in detail in the NAFLD phenotype chapter earlier in the thesis. The first was NAFLD, which was defined by two raised ALT measurements (>19U/L for women and >30U/L for men) at least 3 months apart.[180] The sign up date to GoDARTS was taken as baseline, therefore any patient with 2 raised ALT levels at least 3 months apart before this date was considered a case.

The second phenotype analysed in GoDARTS was the Fatty Liver Index.[148] This combines BMI, waist circumference, GGT and triglycerides using the formula below. This was calculated using measurements for each patient which were taken closest to their sign up date. A full description of FLI is given in a previous chapter.

In the DMDSC cohort, NAFLD was also defined using ALT levels, which is also discussed previously in this thesis. Due to the lack of longitudinal data a single raised ALT rather than two raised measurements three months apart was used to define NAFLD. Fatty Liver Grade

(FLG) was also available for a number of patients in the DMDSC cohort. This was an ordinal scale of fatty infiltration of the liver, from 0-3; none, mild, moderate and severe fatty infiltration respectively.[143] This was measured by abdominal ultrasound, a commonly used and accurate method of detecting steatosis.[30]

4.3.4 Statistical Analyses

Data were analysed using GWAS methodology, in the Plink 1.9 software program.[263] Genome wide significance was considered to be $p < 5x10^{-8}$. Any p values below $5x10^{-6}$ were considered suggestive signals.

Data were analysed separately by cohort, with meta analyses conducted of all cohorts for the NAFLD phenotype. Where NAFLD was the outcome variable, this was analysed in an adjusted binary logistic regression. Adjusted linear regression was used for continuous outcome variables.

Each model was adjusted for a number of covariates, depending on the data available. Ten principle components for the genetic data were calculated for each cohort, and these were included as covariates in the statistical models.

Where many significant results from the same gene were found, LD pruning was performed for the display of the results. Results were pruned with SNPs with Pearson's $R^2 > 0.8$ being omitted from results tables.

The analysis in GoDARTS was adjusted for 10 principal components, age, sex, T2DM and BMI. The individuals in GoDARTS have been genotyped in phases and on different GWAS chips. Individuals genotyped on each chip were analysed together, then the results from the meta-analysed. Analysis of FLI was not adjusted for BMI as the calculation of the FLI score already includes BMI.

In analysis of NAFLD conducted in DMDSC, the logistic regression model was adjusted 10 PCs, sex, age and BMI. All patients in this cohort had T2DM, so this was not included in the model. GWAS of FLG in the DMDSC cohort was adjusted for the same covariates.

Manhattan plots and QQ plots were generated for each GWAS analysis using the "calibrate" and "qqman" R packages.[269–271]

4.4 Results

4.4.1 GoDARTS Data

A GWAS of NAFLD in GoDARTS was run. The Manhattan plot for this analysis is shown below in figure 4-1 with the quantile-quantile (Q-Q) plot for this analysis in figure 4-2.

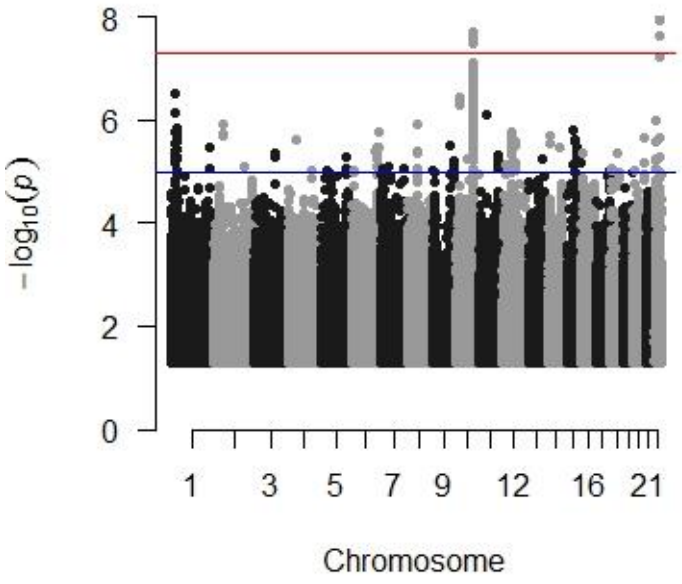

Figure 4-1Manhattan Plot of GoDARTS NAFLD GWAS

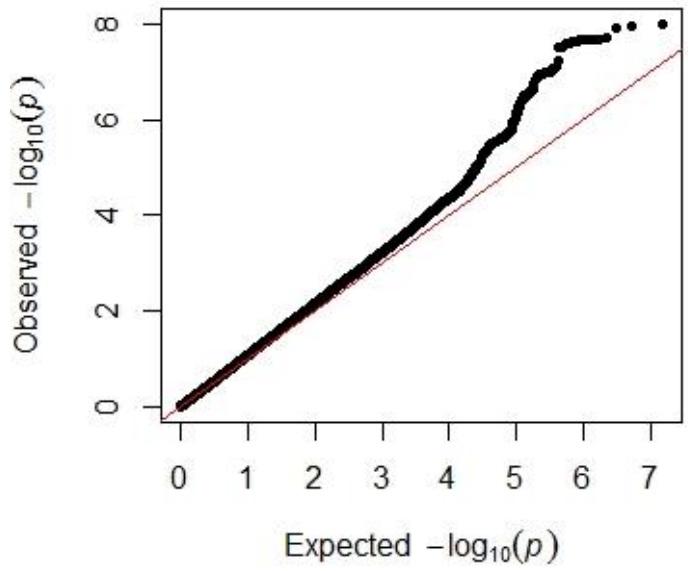

Figure 4-2 Q-Q Plot of GoDARTS NAFLD GWAS

A number of variants in *ERLIN1, BLOC1S2* and *PNPLA3* were genome wide significant for NAFLD, shown in table 10 below. All of the genome wide significant variants in chromosome 10 were in high LD (Pearson's $R^2 > 0.8$), as were the variants in chromosome 22 in *PNPLA3.* (Pearson's $R^2 > 0.9$) An LD pruned table of results is shown below. (Table 10)

Table 11 - Genome wide significant variants for NAFLD in GoDARTS

Chromosome	Gene	Chromosome Position	SNPID	Effect Allele	Effect Allele Frequency	p-value	Odds Ratio	Cochrane's Q	I^2
10	*ERLIN1*	10:101,880,479	rs11594323	A	0.421	3.19E-08	0.816	0.333	11.96
22	*PNPLA3*	22:44,324,727	rs738409	G	0.203	1.09E-08	1.3207	0.2074	36.43

A GWAS of FLI was run in the GoDARTS cohort. The QQ plot and Manhattan plot for this analysis are shown below in figure 4-3 and figure 4-4.

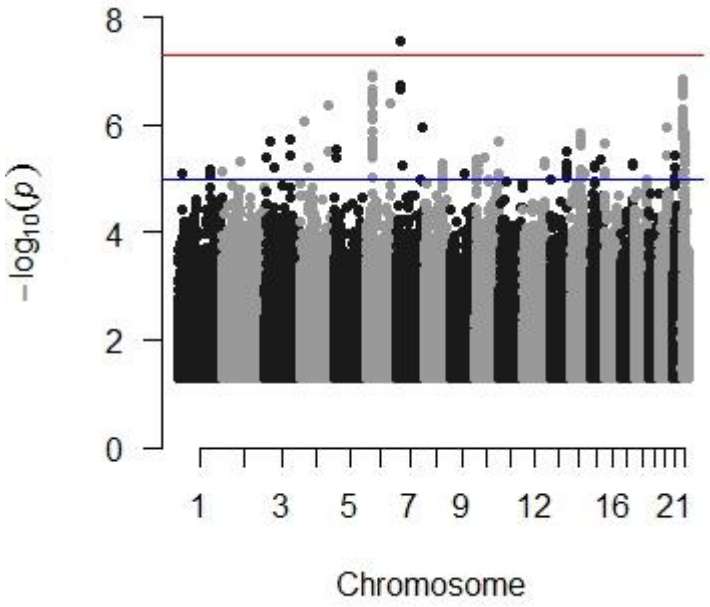

Figure 4-3 Manhattan Plot for GoDARTS FLI GWAS

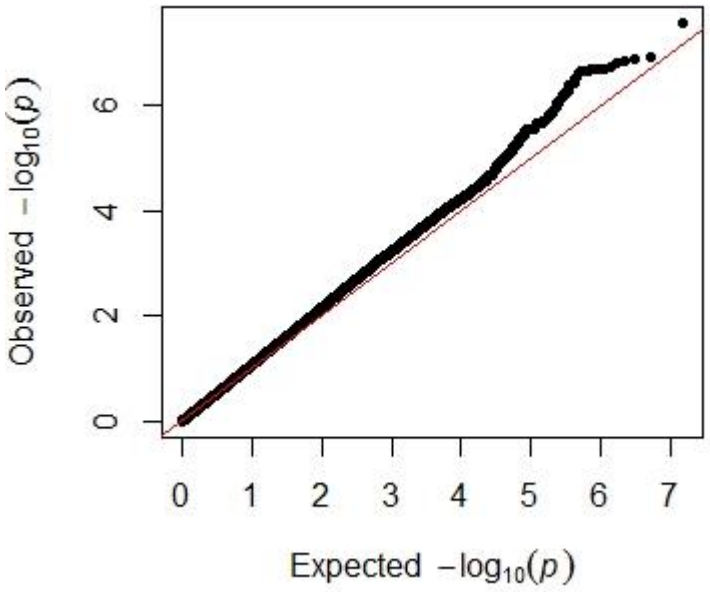

Figure 4-4 Q-Q Plot for GoDARTS FLI GWAS

DNAH11 rs117146188 reached genome wide significance, and several variants reached suggestive significance, shown in table 11 below.

Table 12 Genome wide significant and suggestive significance reaching variants in GWAS of FLI in GoDARTS.

Chromosome	Gene	Chromosome Position	SNPID	Effect Allele	Effect Allele Frequency	P	Beta	Cochrane's Q	I^2
4	*FSTL5*	4:162,824,685	rs190994066	T	0.025	4.19E-07	-7.0671	0.2391	28.83
6	*HLA-DOB*	6:32,756,390	rs114032730	A	0.028	1.17E-07	-6.6811	0.2026	34.95
6	*PSMB9*	6:32,859,137	rs7761882	A	0.030	3.02E-07	-6.5979	0.5054	0
7	*DNAH11*	7:21,883,902	rs117146188	G	0.0138	2.71E-08	-9.3876	0.1968	35.89
22	*GGT1*	22:24,994,708	rs2017188	C	0.0352	1.94E-07	2.3655	0.3577	2.74

A number of variants in *GGT1* were close to genome wide significance for association with increased FLI. These were all in extremely high LD (> 0.99 Pearson's R^2), therefore likely represent one signal. Serum GGT levels have a moderate level of heritability, and some known genetic influences.[272] The *GGT1* variants which were close to genome wide significance in the present study are at a locus which has previously been associated with serum GGT levels.[273] The liver enzyme GGT is a component of the FLI, therefore this *GGT1* locus may represent a confounding variable.[148] This may cause individuals to have higher FLI scores simply because they carry genetic variation in *GGT1*, rather than due to increased levels fatty liver itself.

To overcome this and increase the sensitivity of the analysis, a GWAS of FLI was run again, with *GGT1* rs2017188 included as a covariate. The Manhattan plot and QQ plots for this analysis are shown in figures 4-5 and 4-6 below.

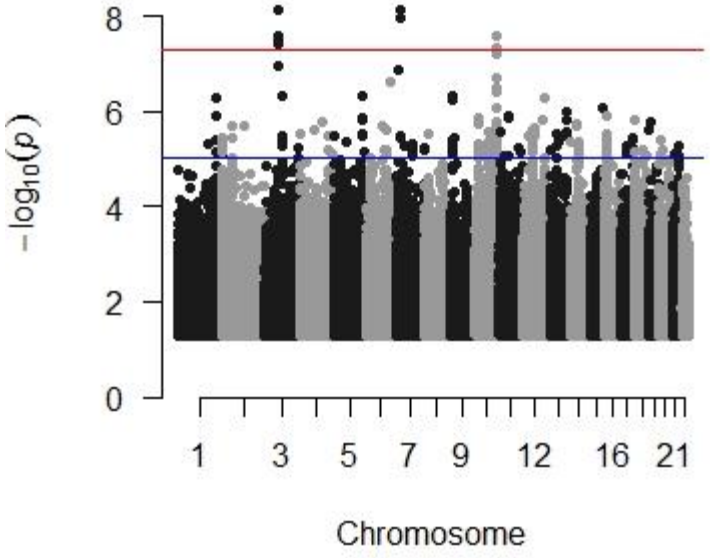

Figure 4-5 Manhattan Plot for GoDARTS FLI GWAS with GGT1 covariate

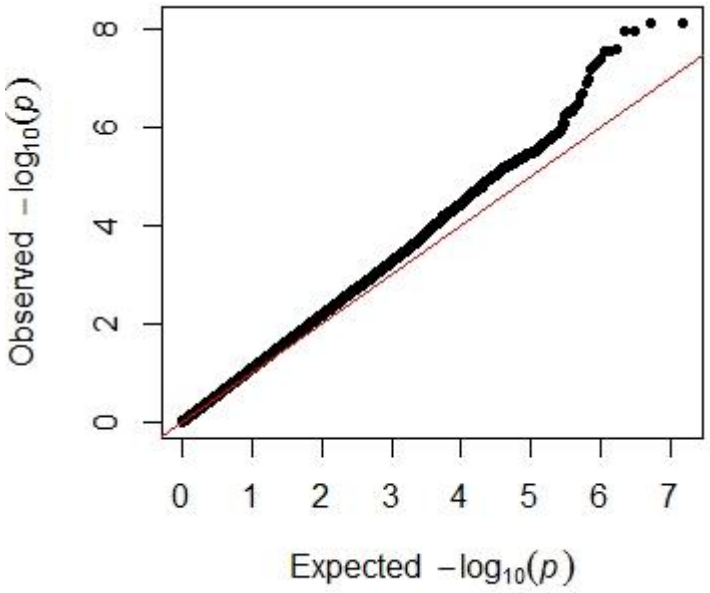

Figure 4-6 Q-Q Plot for GoDARTS FLI GWAS with GGT1 covariate

Genome wide significant results were found in four different genes, shown in table 12 below.

Table 13 - Genome wide significant variants for FLI in GoDARTS

Chromosome	Gene	Chromosome Position	SNPID	Effect Allele	Effect Allele Frequency	p-value	Beta	Cochrane's Q	I^2
3	*FAM19A4*	3:68,998,611	rs1898616	G	0.608	4.00E-08	3.5115	0.369	0
3	*FAM19A4*	3: 69,008,245	rs7427984	A	0.581	2.79E-08	3.4907	0.7684	0
3	*EOGT*	3:69,014,707	rs1506986	A	0.624	7.74E-09	3.7184	0.4439	0
3	*EOGT*	3:69,020,937	rs9853718	G	0.594	2.74E-08	3.5525	0.6323	0
7	*DNAH11*	7:21,883,902	rs117146188	G	0.0138	1.12E-08	-14.0018	0.0786	60.69
7	*DNAH11*	7:21,924,439	rs76307823	C	0.126	1.14E-08	-14.9095	0.3062	15.51
7	*DNAH11*	7:21,930,208	rs77888218	T	0.128	7.70E-09	-15.0827	0.2727	23.04
10	*TCF7L2*	10:114,754,071	rs34872471	C	0.333	2.69E-08	-3.7337	0.1759	42.46
10	*TCF7L2*	10:114,754,784	rs35198068	C	0.334	4.80E-08	-3.662	0.148	47.65

4.5 DMDSC Data

A GWAS was run for the NAFLD phenotype in the DMDSC cohort, with 3,123 eligible participants. The Manhhattan plot for this analysis is shown below in figure 4-7, and the Q-Q

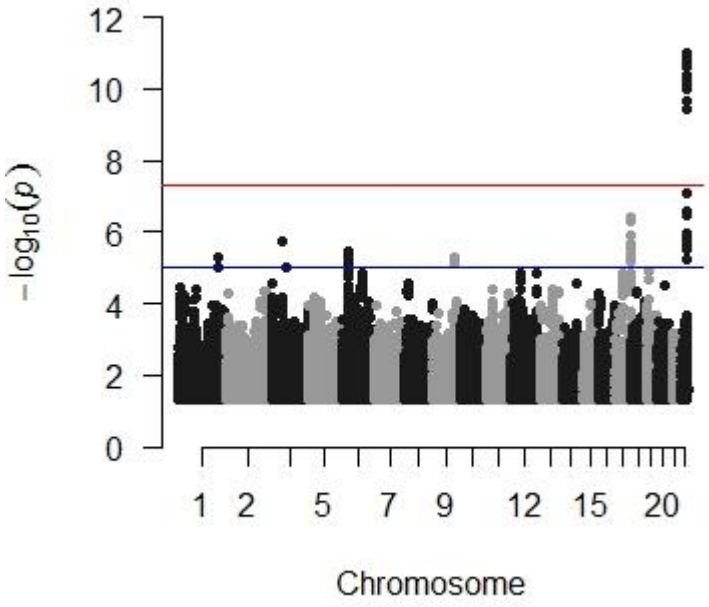

Figure 4-7 Manhattan Plot for DMDSC NAFLD GWAS

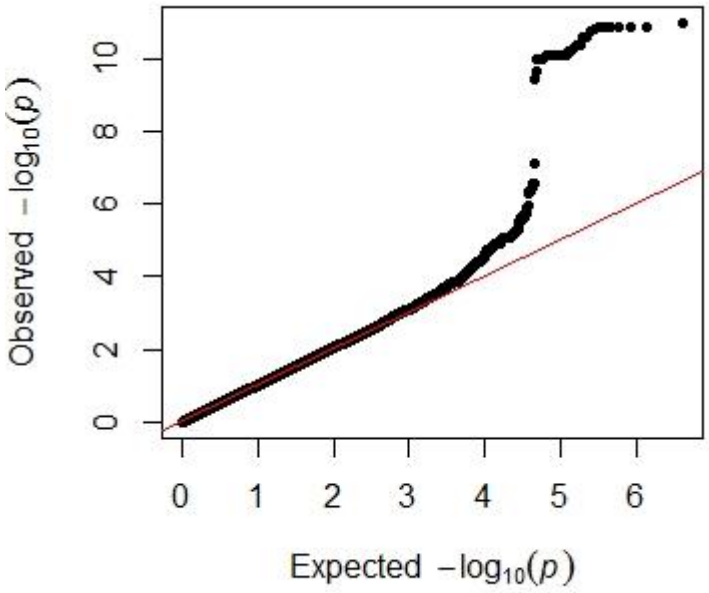

Figure 4-8 Q-Q Plot for DMDSC NAFLD GWAS

plot below in figure 4-8.

A number of variants in *PNPLA3* were genome wide significantly associated with NAFLD, all of which were in very high LD (Pearson's $R^2 > 0.99$). The *PNPLA3* rs738409 variant was among these with an OR of 1.39, and is likely to be the cause of the signal. The results for *PNPLA3* SNPs rs738409 and rs12485100 (which had the lowest p-value out of the SNPs in high LD) are shown in table 13 below.

Table 14- Genome wide significant variants for NAFLD in DMDSC

Chromosome	Gene	Chromosome Position	SNPID	Effect Allele	Effect Allele Frequency	p-value	Odds Ratio
22	*PNPLA3*	22:44,325,516	rs12485100	T	0.242	5.5E-08	1.39
22	*PNPLA3*	22:44,324,727	rs738409	G	0.252	3.0E-08	1.40

Further to the GWAS on NAFLD, a GWAS was run with FLG as the phenotype, with 2,013 eligible participants included in analysis. The Manhattan plot and Q-Q plot for this analysis is shown below in figure 4-9 and 4-10 respectively.

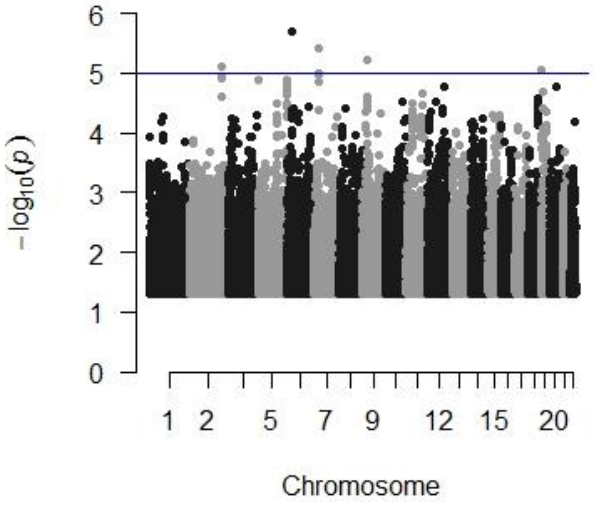

Figure 4-9 Manhattan Plot for DMDSC FLG GWAS

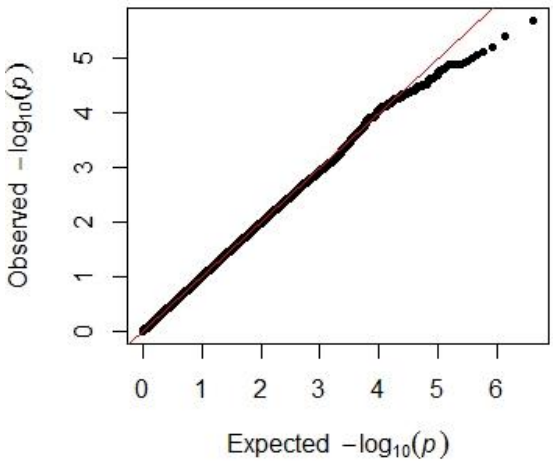

Figure 4-10 Q-Q Plot for DMDSC FLG GWAS

No variants reached genome wide significance for this phenotype, though two variants reached suggestive significance. These are shown in table 14 below.

Table 15 - Suggestive significance reaching SNPs for FLG in DMDSC cohort

Chromosome	Gene	Chromosome Position	SNPID	Effect Allele	Effect Allele Frequency	Beta	SE	p-value
6	*POU5F1*	6:31,148,349	rs114773933	G	0.0176	-0.4205	0.08824	2.02E-06
7	*PDE1C*	7:32,063,068	rs13239020	C	0.0127	-0.48	0.1037	3.92E-06

4.6 Meta Analysis of Scottish and Indian Cohorts

The results of GWAS analyses in the Scottish and Indian cohorts were meta analysed. The results for the meta analysis of NAFLD in the GoDARTS and DMDSC cohorts are displayed in the Manhattan plot and QQ plot in figures 4-11 and 4-12 below.

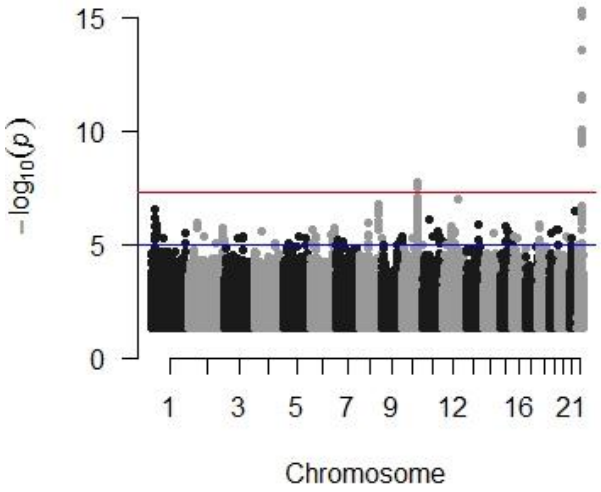

Figure 4-11 Manhattan Plot of DMDSC NAFLD GWAS

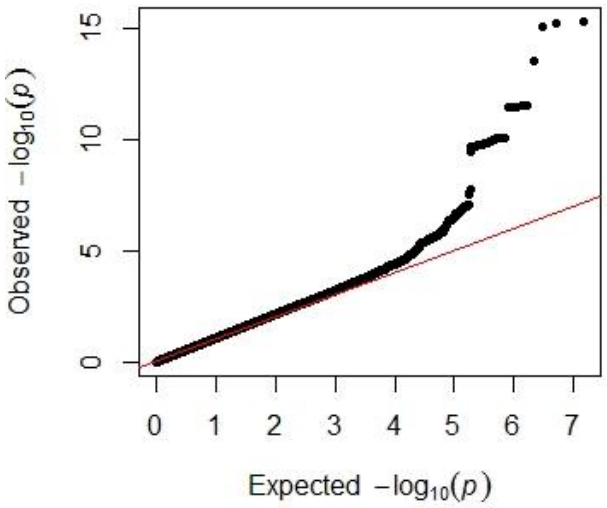

Figure 4-12 Q-Q Plot of DMDSC NAFLD GWAS

Genome wide significant signals were found in *PNPLA3* around the rs738409 locus, the same as was in individual GWAS of GoDARTS and DMDSC. SNPs around *ERLIN1* rs10883447 were also genome wide significant, although these SNPs were not analysed in the DMDSC GWAS as they were excluded during the QC process. The significant variants were pruned for LD greater than 0.8. This shown in table 16 below.

Table 16 - Genome wide significant SNPs for NAFLD in meta analysis of GoDARTS and DMDSC

Chromosome	Gene	Chromosome Position	SNPID	Effect Allele	P	OR	Cochrane's Q	I^2
10	*ERLIN1*	10:101,903,906	rs10883447	G	3.19E-08	0.816	0.333	11.96
22	*PNPLA3*	22:44,324,727	rs738409	G	5.37E-16	1.3423	0.4636	0

A forest plot of the association between *PNPLA3* rs7384098 and NAFLD in the GoDARTS and DMDSC cohorts, as well as the meta analysis, is shown below in figure 4-13.

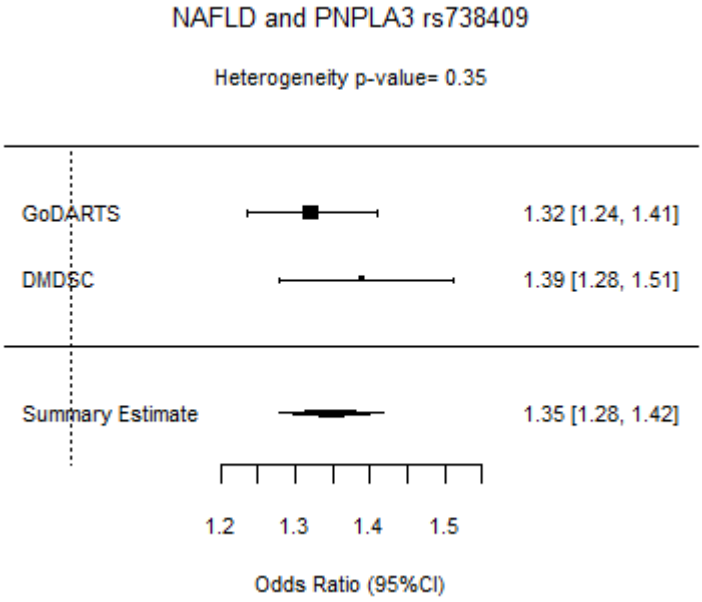

Figure 4-13 Forest plot for PNPLA3 rs738409 in GoDARTS, DMDSC and meta-analysis

Further investigation in *ERLIN1* SNPs in the meta analysis was conducted, as the genome wide significant SNPs were not found in the DMDSC cohort. A number of *ERLIN1* SNPs were analysed both in GoDARTS and DMDSC, though these did not reach genome wide significance. These variants were not nominally associated with NAFLD in DMDSC, although showed the same direction of effect as in GoDARTS. Similarly to GoDARTS, these

were in high LD, with R^2 >0.93. The SNP *ERLIN1* rs11594323 is shown as an example in table 16 below, with results from GoDARTS and DMDSC.

Table 17 - ERLIN1 rs1077821 in GoDARTS and DMDSC NAFLD GWAS

Cohort	Chromosome	Gene	Chromosome Position	SNPID	Effect Allele	P	OR	Cochrane's Q	I^2
GoDARTS	10	*ERLIN1*	10:101,888,520	rs1077821	T	2.096e-08	0.8138	0.3584	6.93
DMDSC	10	*ERLIN1*	10:101,888,520	rs1077821	T	0.5031	0.9577	/	/

A forest plot of the rs1077821 variant in *ERLIN1*, in the GoDARTS, DMDSC and meta-analyses is shown in figure 4-14 below.

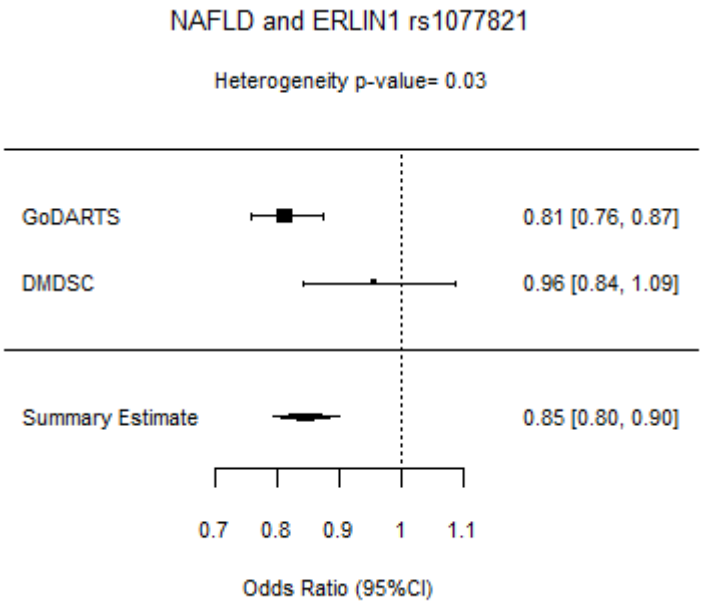

Figure 4-14 Forest Plot for ERLIN rs1077821 in GoDARTS, DMDSC and meta-analysis

4.7 Analysis of Known NAFLD Risk Variants

Following the GWAS analysis, the results were probed for the associations between the relevant phenotype and known NAFLD risk altering variants. Genetic variants with robust and validated associations with NAFLD were selected from previous literature. The

associations between these variants, and the phenotypes analysed in the current study are shown in table 17 below.

Table 18 Associations between known NAFLD risk variants and NAFLD phenotypes in GoDARTS, DMDSC and meta-analysis

		GoDARTS NAFLD		GoDARTS FLI		DMDSC NAFLD		DMDSC FLG		Meta NAFLD	
Gene	SNP	p	OR	p	Beta	p	OR	p	Beta	p	OR
PNPLA3	rs738409	1.09E-08	1.3207	0.2477	0.8816	2.46E-11	1.352	0.0137	0.0663	5.37E-16	1.3423
ERLIN1	rs1077821	2.096E-08	0.8138	0.412	0.524	0.5031	0.9577	0.519	0.01936	2.25E-4	0.90
LYPAL1	rs12137855	0.1195	0.9363	0.93	0.0645	/	/	/	/	0.1195	0.936
HSD17B13	rs6834314	0.2191	0.8522	0.337	-0.671	0.995	0.936	0.918	0.00367	0.237	0.959
TM6SF2	rs58542926	0.3251	1.075	0.843	-0.233	0.16	1.092	0.000194	0.141	0.0405	1.115
NCAN	rs2228603	0.972	1.003	0.176	-1.58	0.0342	1.21	0.001236	0.174	0.229	1.074
PPP1R3B	rs4240624	0.866	1.01	0.18	-2.305	0.241	1.089	0.161	-0.063	0.428	1.027
GCKR	rs780094	0.367	1.035	0.12455	0.998	/	/	/	/	0.367	1.035
FDFT1	rs2645424	0.7544	1.012	0.4713	-0.455	/	/	/	/	0.7544	1.012

A number of known NAFLD risk variants, which were not genome wide significant in the GWAS results, reached nominal significance. ($p < 0.05$) These are highlighted in yellow.

4.8 Discussion

The GWAS analyses in the current study revealed several genetic loci which have significant effects on NAFLD in both GoDARTS and DMDSC cohorts.

4.8.1 NAFLD in GoDARTS

Two genome wide significant variants were found for NAFLD in GoDARTS. These were in *PNPLA3* and *ERLIN1*. Both of these genes have previously been reported as having significant effects on NAFLD risk in GWAS and candidate gene studies.

The *PNPLA3* rs738409 locus showed the strongest association with NAFLD, with an OR of 1.32. This SNP was the first NAFLD risk variant discovered by GWAS; a finding which has been replicated in numerous studies in a variety of populations and ethnicities.[190,274,275] These GWAS and candidate gene studies have found rs738409 to be associated with a range of NAFLD phenotypes, including serum ALT, fibrosis, cirrhosis and hepatocellular carcinoma(HCC).[274,276]Studies have sought to elucidate the mechanism by which *PNPLA3* rs738409 increase NAFLD risk, highlighting several effects.

PNPLA3 is the gene which codes for the protein adiponutrin, which is mainly found in adipocytes and hepatocytes.[277] Its functions include the production and breakdown of fats in hepatocytes.[278] *PNPLA3* rs738409 is a missense variant which interrupts the lipolysis of triglycerides in the liver.[279] It is also associated with decreased release of very-low-density-lipoproteins from the liver.[248] These effects cause disruption in the lipid homeostasis of the liver, increasing the amount of fats stored which becomes steatosis.

The *PNPLA3* rs738409 variant is also associated with increased NASH risk, and histological severity of liver disease.[280] It is also associated with increased risk of fibrosis, cirrhosis and HCC, which has prompted research into whether these associations are driven simply by increased steatosis or another mechanism of *PNPLA3*.[281] In a meta analysis, Singal et al. found that *PNPLA3* increased risk of progression to fibrosis in patients with steatosis regardless of cause.[276] They showed rs738409 increased fibrosis risk in patients with hepatitis C virus (HCV), and others have found increased risk of alcoholic liver disease in those who

carry the rs738409 variant.[282] *PNPLA3* is highly expressed in hepatic stellate cells (HSC), which are the responsible for the production of the extracellular matrix which characterises fibrosis.[283] The rs738409 variant increases the HSC's pro-inflammatory and pro-fibrogenic properties when in an activated state due to hepatocellular injury.[284]

The association between NAFLD and *PNPLA3* rs73809 in GoDARTS is not novel, but acts as a good positive control and validation of the phenotype. An ALT based NAFLD definition is unable to provide any classification of the severity of the disease, but individuals with the full spectrum of NAFLD were included in the study. It is probable that this signal reflects both steatogenic and fibrogenic effects of *PNPLA3,* but simple steatosis in NAFLD is many times more common than the more advanced stages such as fibrosis and cirrhosis.[5] A larger proportion of this signal therefore likely represents increased frequency of steatosis in carriers of the rs738409 variant.[54]

We found that *PNPLA3* rs738409 is associated with NAFLD with an OR of 1.32, in an additive model. This OR is lower than some previously published estimates, some of which are as high as 2.40.[274] The effects of *PNPLA3* rs738409 on NAFLD have been described a number of times in European cohorts.[232,285] These studies have used a number of different NAFLD definitions and related phenotypes however which makes comparison of point estimate of *PNPLA3* effects difficult. Sookoian et al. performed a meta analysis of studies investigating *PNPLA3* rs738409 and NAFLD, finding an overall odds ratio of 3.26 for homozygous wild type versus heterozygotes (CC versus CG).[83] They also found significant association between *PNPLA3* rs738409 and ALT levels across a number of studies. Dai et al. found a lower OR of 2.27 for CC versus CG genotypes in their analysis.[275] The inclusion criteria for this study required that NAFLD be diagnosed by MRI, ultrasound or liver biopsy, which are more accurate than LFTs for NAFLD diagnosis, but are predominantly performed only when liver disease is suspected.[30,135,286] This means that asymptomatic cases are likely to

be untested and therefore misclassified. The differences in phenotypes may explain some of the variability in odds ratios.

Variants in *ERLIN1* were found to reduce NAFLD risk with odds ratios of ~0.81 in GoDARTS. *ERLIN1 (Endoplasmic Reticulum lipid raft protein 1)* encodes a prohibitin protein that defines lipid-raft-like domains of the endoplasmic reticulum.[287] Lipid rafts are structures made of lipids and proteins which are represent in plasma membranes, in this case of the endoplasmic reticulum(ER). Disordering of the ER lipid raft can disrupt the function of the ER, including cell signalling.[288] ER function is a key factor in NAFLD, as ER stress can contribute to several mechanisms which increase hepatic steatosis, fibrosis and cell death.[289]

The protein encoded by *ERLIN1* binds to cholesterol and regulates the SREBP signalling pathway.[290] ERLIN proteins restrict the release of SREBP from the ER, and it has been shown that in the absence of ERLIN proteins, SREBP activity is increased.[290] SREBPs are transcription factors which bind to sterol regulatory elements and upregulate them to produce more enzymes which are required for sterol biosynthesis.[291,292] SREBPs are key regulators of cell lipid homeostasis, involved in a number of processes including; global lipid synthesis and growth, fatty acid synthesis, energy storage and cholesterol regulation.[293] SREBPs are associated with NAFLD through several pathways, increasing steatosis and inflammation, partially through increased ER stress.[294] Genetic variants in the genes which encode SREBPs have been linked to NAFLD.[295]

ERLIN1 is part of a cluster involving the genes *CHUK-CWF19L1-ERLIN1*, and a number of variants in these genes are in LD. The variants in *CHUK* and *CWF19L1* form one haplotype block in very high LD (>0.99), and the variants in *ERLIN* another haplotype block with high LD with each other (>0.82 -0.99). These two haplotype blocks are also in high LD, with Pearson's R^2 ~0.80 between variants in each block. This cluster has been linked to NAFLD

previously, with associations with ALT, NAFLD defined by CT scan and NASH.[287,296,297] Three variants in ERLIN1; rs10883451, rs1408579 and rs2862954; were found to be genome wide significantly associated in a meta-analysis of ALT and NAFLD.[296]

It is unclear whether both of these gene haplotype blocks have independent effects on NAFLD, or if just one of the genes is the causative source of the association. No previously published studies have reported analysis of this. The *CHUK* gene encodes a protein kinase which inhibits the essential transcription nuclear factor-kappa-beta (NF-κB) complex.[131] NF-κB is associated with the regulation of cell activities including inflammation and cell death.[298] Given what is known about the function of each gene, it is plausible that both *CHUK* and *ERLIN1* have independent effects on NAFLD, although further research is required to characterise their relationship to NAFLD clearly.

Furthermore, there is evidence that these variants in *ERLIN1* are expression trait quantitative loci (eQTL).[299] *ERLIN1* is a cis-eQTL for, a neighbouring gene.[296] *CWF19-like 1 cell cycle control (CWF19L1)* variants were found to be associated with NAFLD and ALT levels in the study by Feitosa et al.[296] In the UK BioBank cohort, the *CWF19L1* rs17729876 variant is associated with T2DM, and chronic liver disease.[300] Data from The Human Protein Atlas show *CWF19L1* is associated with increased liver cancer mortality(http://www.proteinatlas.org).[301] This locus is also linked to increased cholesterol levels.[302] These findings may suggest that the association between *ERLIN1* and NAFLD is due through expression of *CWF19L1*.

4.8.2 FLI in GoDARTS

A GWAS analysis of Fatty Liver Index in GoDARTS found one genome wide significant signal, *DNAH11* rs117146188 and showed a number of variants which were close to genome wide significance. Ten of these variants were in the *GGT1* gene, in high LD. ($R^2 > 0.99$) This

result was germane as gamma-glutamyltransferase (GGT) is a biomarker which is used in the calculation of FLI. The *GGT1* gene encodes GGT, and variants in this gene have been reported to influence serum GGT levels. The *GGT1* variant rs2006227 is among those reported to influence serum GGT levels, and was among the ten *GGT1* SNPs which were close to genome wide significance in the current study.[303] It is therefore likely that the association between *GGT1* variants and FLI is driven by the association with serum GGT levels, and not by altered NAFLD risk.

To control for this genetic risk factor for GGT, and thus higher FLI, a second GWAS of FLI was run in GoDARTS with *GGT* rs2006227 included as a covariate. This produced stronger results, and a number of genome wide significant hits.

DNAH11 rs117146188 was significantly associated with FLI in both analyses, adjusted and non-adjusted for *GGT1*. Two more variants in *DNAH11* were also genome wide significantly associated with FLI; rs76307823 and rs77888218; which were in almost perfect LD (R^2 = 0.99). These two variants were in high LD with rs117146188, with a Pearson's R^2 of 0.87. The β estimate for rs117146188 was -14.0 per allele, a strong protective effect against high FLI. This effect size is large, given that FLI extends from 0 to 100.[148] *DNAH11 (Dynein Aaxonemal Heavy Chain 11)* encodes a ciliary outer dynein arm protein, and is involved in the movement of respiratory cilia.[304] It has not previously been linked to NAFLD or related conditions.

In the analysis adjusted for *GGT1* three more variants were significantly associated with FLI; *FAM19A4*, *EOGT* and *TCF7L2*. Two variants in *FAM19A4 (Family with Sequence Similarity 1 member A4)* which were in high LD (R^2= 0.93) were genome wide significantly associated with FLI; rs1898616 and rs7427984. The function of *FAM19A4* is not clear, though it has been found to have a role in regulation of macrophages in response to inflammation.[305]

FAM19A4 is an eQTL of *EOGT,* which may explain the association seen between *FAM19A4* and FLI.[306]

Two variants in *EOGT* (rs1506986 and rs9853718) were significantly associated with FLI in the GWAS adjusted for *GGT1*, and were in high LD.($R^2 > 0.99$) *EOGT(EGF Domain Specific O-Linked N-Acetylglucosamine Transferase)* encodes a protein which is active in the ER of the cell and catalyses the transfer of N-acetylglucosamine (GlcNAc) to extracellular proteins.[307] The adding of GlcNAc (O-GlcNAcylation) activates or deactivates enzymes and transcription factors, and is a mechanism of regulation. In the liver, this process is a key factor in regulation of metabolism.[308] O-GlcNAcylation is associated with hepatic insulin resistance and disruption of a number of processes in the liver, including gluconeogenesis, glycolysis and glycogenesis. [309,310] Studies have suggested that this is in some part through regulation of the *FXR* gene, which regulates *SREBP-1c,* a key regulator of hepatic lipid homeostasis.[293] The variants in *EOGT* which were significant in the current study have no reported associations with clinical outcomes in published literature, though data from the IEU OpenGWAS system shows *EOGT* rs1506986 is associated with microalbumin in urine and eosinophil count.[311] In the UK Biobank data, *EOGT* rs9853718 is associated with abdominal aortic aneurysm.[300,312]

TCF7L2 (*Transcription Factor-7–Like 2*) rs34872471 and rs35198068 were both genome wide significantly associated with FLI in the GWAS adjusted for *GGT1.* These variants are in high LD. ($R^2 > 0.99$) *TCF7L2* influences the transcription of a number of genes, and notably regulates glucose metabolism in the pancreas and liver.[313] rs34872471 is associated with increased T2DM risk, which has been shown in a number of populations.[314,315] It is also associated with decreased blood pressure.[312] This locus is in almost perfect LD with the common T2DM risk variant rs7903146 ($R^2 > 0.99$), which likely explains these associations. This variant increased diabetes risk by altering incretin action, and is associated with

cardiovascular disease.[316,317] It has been reported to increase NAFLD risk independently of diabetes in several studies.[254,318] The current study concurs with these results, and confirms that there is increases steatosis in carriers of the *TCF7L2* rs7903146 variant.

4.8.3 NAFLD in DMDSC

Variants in the *PNPLA3* gene around the rs738409 locus were genome wide significantly associated with NAFLD. All significant variants were in almost perfect LD ($R^2 > 0.99$). The OR for NAFLD was 1.39 for rs738409. This variant has previously been reported to be a risk factor for NAFLD in Indians, as well as South Indians specifically.[255,256,319] Bale et al. found that *PNPLA3* was associated with NAFLD risk for North Indians compared to South Indians.[256] The findings of the current study confirm the relationship between *PNPLA3* rs738409 and NAFLD, and act as a positive control, validating the NAFLD phenotype.

4.8.4 FLG in DMDSC

No variants were genome wide significantly associated with Fatty Liver Grade in the DMDSC cohort. Two variants however reached suggestive significance. *POU5F1* rs114773933 showed suggestive significance for negative association with FLG. *POU5F1 (POU Domain, Class 5, Transcription Factor 1)* encodes a transcription factor protein which regulates cell differentiation.[320] This variant has not been linked to clinical outcomes in any published literature. This pathway has been linked to NASH, as Chien et al. showed that POU5F1 could improve effectiveness transplanting pluripotent stem cells to treat both steatosis and steatohepatitis.[321] Park et al. showed similar results, with stem cells induced by POUF51 and HNF1α improved chronic liver injury.[322] It has also been implicated in liver cancer, as POU5F1 drives self-renewal of liver cancer cells.[323] The current study suggests a role of the POUF51 transcription factor and a variant of its associated gene in the development of steatosis.

PDE1C rs13239020 was negatively associated with FLG, with a suggestive p value. *PDEC1 (Phosphodiesterase 1C)* encodes an enzyme regulates proliferation and migration of vascular smooth muscle cells.[324] The PDE1C enzyme downregulates glucose dependent insulin secretion, and inhibition of PDE1C upregulated insulin secretion.[324] Patients with NAFLD have impaired incretin effect.[325] The association between *PDE1C* and FLG in the current study may be a result of this alteration of insulin secretion.

4.8.5 Meta-Analysis

The *PNPLA3* variant rs738409 was found to be genome wide significantly associated with NAFLD in both the GoDARTS and DMDSC cohorts, as well as the meta analysis. Both had similar odds ratios with GoDARTS = 1.32 and DMDSC = 1.39, and both European and South Asian populations are reported to have a MAF of ~0.22 for this variant.[326]

The association between this locus and NAFLD has been demonstrated in both European and South Indian populations in previous literature.[42,327] Previous literature has found that there are ethnic differences in the magnitude of the NAFLD risk associated with *PNPLA3* rs738409. In a meta analysis of 13 articles, Dai et al. showed that this variant had a larger effect in Caucasian individuals compared to Asian individuals. However, the Asian cohorts meta analysed in this study were predominantly Chinese, and may be different to Indian populations with respect to *PNPLA3*. Gnomad reports differences in MAF for rs738409, with MAF = 0.3816 in East Asians and MAF = 0.22 in South Asians.[326] The estimate for Asian individuals in the Dai et al study therefore may not be applicable to the DMDSC cohort. Due to the differences in NAFLD case ascertainment, and overall differences in the data, we are unable to draw any conclusions about the relative magnitude of the *PNPLA3* rs738409 effect in each of the cohorts.

The *ERLIN1* locus near rs10883447 was significantly associated with NAFLD in GoDARTS, but were not significantly associated with NAFLD in DMDSC. Though *ERLIN1* rs10883447 was excluded in the QC process in DMDSC due to missingness > 0.02, some *ERLIN1* variants in high LD were analysed. *ERLIN1* rs1077821 showed the same, protective direction of effect as rs10883447 did in GoDARTS. Significant heterogeneity was found between the results for rs1077821 in DMDSC and GoDARTS, with an I^2 value of 79.37%.[328] This statistic may suggest a difference in the effect of *ERLIN1* between the cohorts, though this could be due to lack of experimental power. Variance at this locus is rare in the DMDSC cohort compared with GoDARTS, as for rs1077821, the MAFs are 0.22(DMDSC) and 0.42 (GoDARTS). Similar differences in frequency between Europeans and South Asians are also reported in previous studies such as 1000 Genomes.[268] The relative rarity of the effect allele combined with the lower number of individuals analysed (N = 3,133), could mean that the analysis in DMDSC was not adequately powered to detect an association. The experimental power was calculated with an alpha level of 0.05, giving a power of 3.1%.[329]

4.8.6 Known NAFLD Variants Analysis

Following the GWAS analysis the association results for a number of known NAFLD risk variants were compiled. These variants were selected based on their identification as NAFLD risk variants in previous literature. In the GoDARTS NAFLD phenotype, aside from the genome wide significant signals in *PNPLA3* and *ERLIN1* none of the tested variants reached nominal significance. Further investigation was conducted to find why many known NAFLD variants were not significantly associated with NAFLD in GoDARTS was undertaken. It was found none of the variants had significant heterogeneity of effect between individuals genotyped on each of the three platforms used. This non-significance of many SNPs may be a reflection of the phenotypes chosen. Certain NAFLD variants are associated with particular features and stages of NAFLD, and these may not have been picked up by our phenotype.

Failed replication of genetic studies is common, as Wu et al. found no association between NAFLD and *NCAN* rs2228603 in a Han Chinese population, where others had shown an association in previous studies.[330]

Issues of statistical power may have altered the chances of seeing significant replication results. In GoDARTS for example, for the *TM6SF2* rs58542926 variant which has a MAF of 0.068, the statistical power for the analysis was 56.2% with an alpha level of 0.05. This power level is below threshold of 80%, which is the most commonly used level in medical studies.[331]

Further to this an additive model was used, which is not the model used in the discovery of all of these variants. This is applicable to many of the variants tested, but this may have underestimated the effects of a number of SNPs as they behave in dominant, recessive or overdominant ways rather than additively.[332,333] Other models were tested for a number of variants and it was found that *NCAN* rs2228603 was significantly associated with NAFLD in a recessive model in GoDARTS. (OR =2.08(1.08 - 4.16), p = 0.031)

GWAS analyses reported in previous literature have also included different covariates in their analyses, and this may have had an influence on the lack of replication in the current GWAS.[252,334] For example, the NAFLD risk variant *TM6SF2* rs58542926 was not significantly associated with NAFLD in the GWAS in GoDARTS, but was when run in an unadjusted model. (1.13(1.01 - 1.26), p = 0.035) These factors, as well as differences in phenotype, data collection, cohort, ethnicity and genotyping platform, can make the interpretation of negative results in GWAS analyses challenging.

In the DMDSC cohort, a number of known NAFLD variants were associated with NAFLD. *PNPLA3* rs738409 was found to be genome wide significantly associated with NAFLD, though *ERLIN1* variants were not tested in this analysis as they were excluded during the QC

process. The *NCAN* variant rs2228603 was associated with NAFLD with an OR = 1.21. *NCAN* is involved in cell adhesion and migration in the nervous system, which is increasingly thought to play a role in metabolism.[335] This variant has been reported to increase NAFLD risk in a number of studies primarily among Europeans.[335,336] A study in a Han Chinese study failed to replicate this finding.[330] A study in an Indian population found associations between this variant and ALT levels, as well as NAFLD although this was not adjusted for multiple testing.[100] The current study confirms the association with ALT levels, and suggests further that there may be an association with NAFLD.

A number of known NAFLD variants were nominally associated with FLG in the DMDSC cohort. *PNPLA3* rs734409 was associated with increased FLG, with a β estimate of 0.0663. Though the effect size was modest, the same direction of effect for this variant on NAFLD and related phenotypes is present.

TM6SF2 rs58542926 was significantly associated with increased FLG, with a β of 0.141. *TM6SF2 (Transmembrane 6 Superfamily Member 2)* is a key regulator of liver fat metabolism.[337] This variant has been linked with NAFLD and NAFLD progression to NASH in a number of studies.[191,338,339] The *NCAN* variant rs2228603 was also associated with increased FLG in the current study.

In the meta-analysis of NAFLD between DMDSC and GoDARTS data, other than the genome wide significant variants in *PNPLA3* and *ERLIN1,* none of the tested variants reached nominal significance.

4.8.7 Limitations and Comparability of Results

The definition of NAFLD in GoDARTS and DMDSC may be a weakness of the current study. This is discussed in depth in a previous chapter of this thesis outlining the NAFLD

definition. Despite some patients presenting with NAFLD without raised ALTs, we demonstrated that the NAFLD definition in GoDARTS is sensitive and reliable.[340]

The Fatty Liver Grade phenotype in the DMDSC cohort was based on abdominal ultrasound scans, a non-invasive and accurate method of diagnosis.[196,341] It has been shown that grading of steatosis level can predict the histologic severity of liver disease with some accuracy, but intra-observer variability can affect the reliability of this phenotype.[342] This issue is apparent in all measures which require assessment from an observer, and are not fully quantitative. It was shown in an earlier chapter that the FLG phenotype correlated well with ALT, as well as other known metabolic and anthropomorphic features such as HbA1c, BMI and Waist. In the analysis of known NAFLD variants, SNPs from *PNPLA3, NCAN* and *TM6SF2* were nominally associated with FLG, which is further validation of the accuracy of this phenotype.

The NAFLD phenotype based on raised ALT levels in the DMDSC cohort was less specific than the NAFLD phenotype in GoDARTS due to the unavailability of certain data. This NAFLD phenotype is developed and discussed in depth in a previous chapter, along with some of its limitations. In DMDSC, data about alcohol intake is unavailable, which increases risk of alcoholic liver disease (ALD) cases being classified as NAFLD. In GoDARTS, access to EHRs for enrolled participants allowed the exclusion of other causes of liver disease, such as immunological or viral insults. This was not possible in the DMDSC cohort, meaning that the NAFLD and FLG phenotypes likely included patients with non-NAFLD related liver disease as cases.

Comparison of results may yield insights about the drivers of NAFLD in each population. However, the GWAS studies run in the GoDARTS and DMDSC cohorts have a number of differences which make the direct comparison of results challenging. Differences in the data availability and source mean that comparing the variants discovered by GWAS between each

cohort is not a reliable way of determining differences in which variants affect NAFLD in each cohort.

There were large inconsistencies in the amount of data available for individuals in each cohort. In the GoDARTS cohort EHRs from the NHS were used as the source of data. This gave us access to longitudinal measurements for ALT, allowing a NAFLD definition with increased specificity and sensitivity. The DMDSC data had ALT measurements from clinic visits, and the NAFLD definition was based on a single measurement taken at the beginning of the study period. The heterogeneity between these phenotypes could cause differences in the variants which are found to be significant by the GWAS, which reduces the validity of direct comparison.

Related to the differences in the data available, is the source of data and way in which it was collected. These were longitudinal datasets and contained many years' worth of measurements for each patient, pre and post study commencement. The DMDSC data on the other hand was from EHRs from private clinic visits. Individuals had measurements taken on their first and subsequent visits to the clinics, most of which were at time of T2DM diagnosis. This disparity in the way the data was collected could have affected the NAFLD phenotype and therefore the results of the GWAS.

The GoDARTS cohort comprises both T2DM patients and healthy controls whereas the DMDSC cohort is solely individuals with T2DM. This is another factor which may impact the prevalence and presentation of NAFLD in the cohort under analysis, thus altering results and inferences drawn from them.

The analysis of known NAFLD variants in the GWAS results had some issues which may have caused many of the non-significant results for each SNP. A number of these issues are

discussed above some of these issues are discussed above, and outline why interpretation of negative results from GWAS can have limited usefulness due to these factors.

4.9 Conclusion

A number of genetic loci which influence NAFLD risk were found in GWAS analyses. PNPLA3 rs738409 was a significant risk factor for NAFLD in both the Scottish and Indian cohorts, with similar effect size and similar minor allele frequency. This demonstrates that both populations share a common genetic risk factor, and that the NAFLD seen in Scotland and India is in many ways the same. This combined with previous research demonstrate that PNPLA3 rs738409 is a key NAFLD risk variant in populations across the world. The ERLIN1 locus around the missense variant rs10883447 was also associated with increased NAFLD risk in the Scottish population, but not the Indian population. Differences in phenotype, as well as lack of statistical power are likely to have contributed to this difference between cohorts.

5 *GLP1R, GCG* and *GCGR* Genes and NAFLD

5.1 Abstract

Dual agonist medications for glucagon-like peptide-1 receptor/glucagon receptor (GLP-1R/GCGR) have shown promising results in treatment of obesity and type 2 diabetes mellitus (T2DM). Given the overlap in pathophysiology and epidemiology of these conditions and non-alcoholic fatty liver disease (NAFLD), dual agonists are being considered for treatment of NAFLD. The aim of this study was to investigate the effects of *GLP1R,* GCG and *GCGR* genetic variants on NAFLD rate.

Analyses for this cohort study were conducted in the GoDARTS and SHARE cohorts, two Scottish cohorts of 13,695, and 62,438 individuals respectively. Meta-analysis of these cohorts was also conducted. Further replication was conducted in the DMDSC cohort, consisting of 3,154 South Indian individuals with T2DM. The NAFLD phenotype was defined as at least 2 elevated ALT measurements recorded at least 3 months apart. Common variants (>1% MAF) from *GLP1R, GCG* and *GCGR* were selected for analysis, some of which are known to affect T2DM and metabolic factors.

Two variants from *GLP1R* were associated with NAFLD rate. In the meta-analysis, rs6923761 recessively increased NAFLD risk in a model adjusted for sex, age and T2DM. (OR = 1.15(1.01 - 1.31), p = 0.032) Another missense variant in *GLP1R*, rs1042044, decreased NAFLD risk. (OR = 0.88, p = 0.018) The *GCGR* variant rs140065949 was associated with increased NAFLD. (OR = 1.40, p = 0.029) A number of statistically significant gene/gene interactions were found.

These findings demonstrate that *GLP1R* and *GCGR* variants are associated with NAFLD risk. This combined with previous literature on these SNPs support the notion that co-agonism for *GLP1R* and *GCGR* may be effective treating NAFLD.

5.2 Introduction

Non-alcoholic fatty liver disease (NAFLD) is the most common cause of liver disease globally, affecting around 25.2% of adults worldwide.[1] It is a prevalent comorbidity of obesity, and frequently occurs in individuals with and type 2 diabetes (T2DM).[343] There is currently no specific pharmacological intervention recommended for NAFLD.

Dual agonist medications for glucagon-like peptide-1 receptor/glucagon receptor (GLP-1R/GCGR) have had promising results in treatment of obesity and T2DM.[344] GLP-1 is an incretin which increases secretion of insulin, lowering blood glucose levels.[345] Glucagon, conversely, raises blood glucose levels; stimulating the liver to convert glycogen to glucose and release this into the bloodstream.[346] Despite this, glucagon has thermogenic and catabolic effects which are desirable for the treatment of obesity and diabetes, and is therefore a promising treatment target.[344] Another desirable effect of glucagon is the reduction in food intake.

These effects are combined into GLP1/GCGR co-agonists. These upregulate GLP1 and GCG receptors, and have been shown to reduce obesity, and enhance insulin secretion.[347] Although some of the effects of these receptors are diametrically opposed with regards to blood glucose levels, it appears that the increased activation counteracts the undesirable gluconeogenesis and glycogenolysis stimulating effects of glucagon.[344]

[344] The effects of increased GCGR and GLP1R activation, as well as the effects of GLP1R/GCGR coagonists are shown in figure 5-1 below.

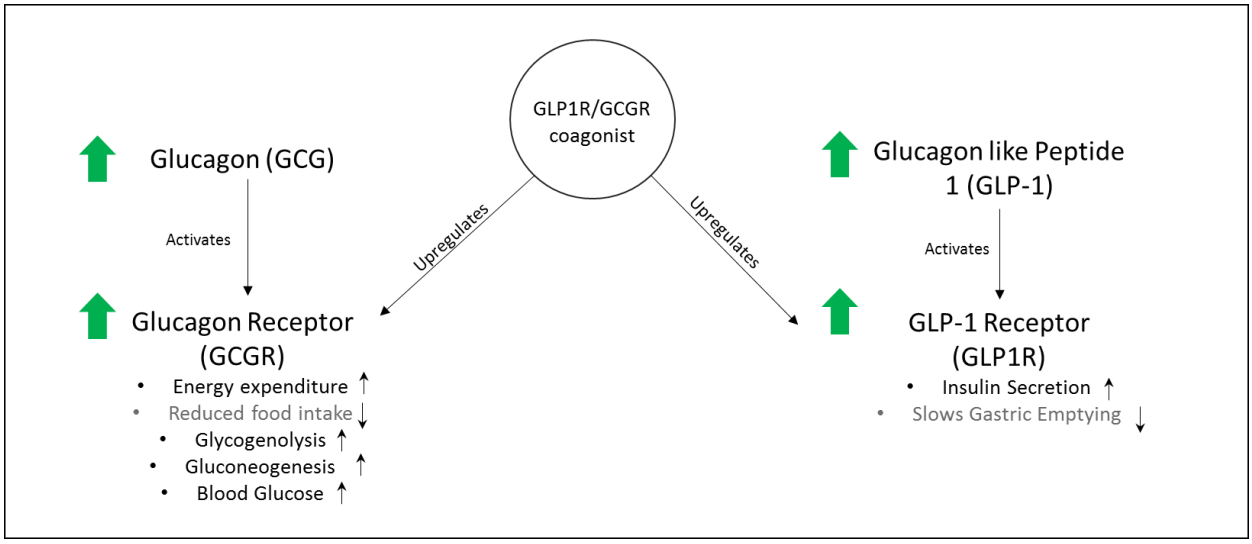

Figure 5-1Effects of increased GLP1R and GCGR activation, and role of GLP1R/GCGR coagonists.

Given the overlap in pathophysiology between obesity and T2DM, and NAFLD, dual agonist medications for GLP1R and GCGR have been investigated as a potential therapy for NAFLD. Preliminary studies have shown a reduction in hepatic steatosis and inflammation as a result of GLP1R/GCGR co-agonists.[97] It is thought that the action of GLP1R and GCGR together will decrease the amount of triglycerides stored in the liver, and improve hepatic insulin resistance.

Investigation of genetic variants associated with the mechanisms by which diseases and medications work can reveal important information. This can be used to develop medications, and to stratify individuals. GLP1R and GCGR each have a gene which codes for them.[348,349] Variations in these can affect the function of the receptors; for example rare defects in *GCGR* can cause non-insulin-dependent diabetes mellitus.[349] Variants in *GLP1R* have been found to influence drug response.[350]

The aim of this study was to investigate the effects of common genetic variants in *GLP1R, GCG and GCGR* on NAFLD. This will aid the investigation into GLP1R/GCGR co-agonists

for NAFLD, and help achieve personalised medicine through genetic stratification for individuals if these co-agonists become a mainstream therapy for NAFLD.

5.3 Methods

The GoDARTS study was used as a discovery cohort.[187] This is a T2DM case-control cohort with electronic health record data available for the 18,306 participants and genetic data available for many of these. Of these individuals, 10,021 had at least one of the variants of interest genotyped and were suitable for analysis once exclusions for alternate causes of liver disease were made.

The analyses were also conducted in the SHARE cohort for validation, and meta-analysis of results.[188] This cohort comprised 73,024 individuals, 3,068 of whom had been genotyped and were suitable analysis.

Supplementary validation was conducted T2DM cohort from Dr Mohan's Diabetes Specialities Centre (DMDSC).[195] This cohort comprised 75,952 individuals who were patients of the DMDSC and had T2DM. These individuals were predominantly South Asian. From this cohort, there were 3,154 individuals who had been genotyped and had sufficient data available for analysis.

5.3.1 Measures

The main outcome measure of this study was NAFLD. This phenotype was defined by raised alanine transaminase levels, a simple and reliable biomarker for NAFLD.[31,158] Any serum ALT measurement greater than 30 U/L for men and greater than 19 U/L per litre for women was considered elevated, based on the values suggested by Prati et al.[180]

In the DMDSC cohort, longitudinal measures of ALT levels were not available, so a single raised ALT measure was used to define NAFLD. This was measured at patients' first visit to the DMDSC T2DM clinic.

5.3.2 Exclusions

To ensure the specificity of the NAFLD definition individuals with features of alternate causes of liver disease were excluded. any positive serological tests for anti-smooth muscle antibody, antinuclear antibodies or anti-mitochondrial antibodies, any positive serology for hepatitis B surface antigen or hepatitis C antibody, or mention of cause of liver disease in medical records. Individuals with alcohol dependence or any documentation of alcoholic liver disease in their EHRs were excluded. In addition, individuals who self-reported drinking more than 20g a day for women and more than 30g a day for men were excluded.

In the DMDSC cohort, medical record data which could be used to rule out those with alterative causes of liver disease was not available. A data field noting whether the patient consumed alcohol ever or never was available, so this variable was included as a covariate in all models.

5.3.3 Genetic Variants Analysed

The genetic variants analysed in this study are shown in table 18. Global MAFs were taken from gnomAD v2.1.1.[351]

Table 19- Frequencies and functions of GLP1R and GCG, GCGR SNPs analysed in the current study

Gene	SNP	GoDARTS MAF	SHARE MAF	Global MAF	Effect	Description	Reference Allele	Alternate Allele
GLP1R	rs6923761	0.359	0.349	0.229	p.Gly168Ser	Missense	G	A
GLP1R	rs1042044	0.582	0.578	0.565	p.Leu260Phe	Missense	A	C
GLP1R	rs10305420	0.407	0.391	0.307	p.Pro7Leu	Missense	C	T
GCGR	rs140065949	0.017	0.027	0.027		Intron	C	T
GCG	rs4664447	0.022	0.021	0.036		Intron	T	C
GCGR	rs28454947	0.161	0.196	0.175	p.Gly229Gly	Synonymous	T	C
GCGR	rs5386	0.090	0.095	0.070	p.Ala155Ala	Synonymous	C	G

These were three common missense variants in *GLP1R*, plus three common variants in *GCGR* and one in *GCG*. Variants with a minor allele frequency MAF of over 1% were selected. Three of these variants (rs28454947, rs5384, rs2272030) were found to be in LD ($R^2 > 0.9$), therefore only the most common - rs28454947 - was analysed from these. The correlation matrix of the SNPs in *GLP1R, GCG* and *GCGR* is shown in figure 5-1 below.

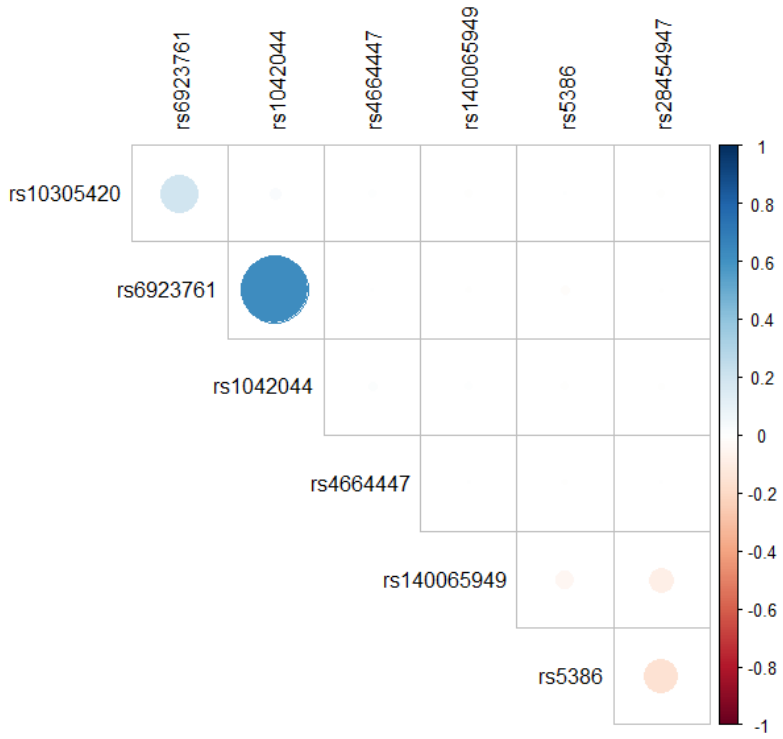

Figure 5-2 Correlation Matrix of Candidates SNPs

5.3.4 Analysis Methods

The association between the genetic variants of interest and NAFLD was assessed in a logistic regression model. They were tested in a number of unadjusted and unadjusted models. The adjusted models contained age, sex, and T2DM as covariates. A secondary adjusted model included these variables, plus BMI also. A large number of individuals in the SHARE cohort were missing BMI data, therefore models with and without BMI as a covariate were used. These models were run in GoDARTS, SHARE and then an individual participant data (IPD) meta-analysis of GoDARTS and SHARE cohorts.

There were several sources of heterogeneity between the Scottish cohorts, and the DMDSC cohort, which led to this cohort not being included in the meta-analysis. The main difference was the NAFLD phenotype, which was based on a single ALT measurement rather than 2 which were 3 months apart. Other differences in the data availability and the way in which it was collected mean that it was not suitable for meta analysis.

5.4 Results

5.4.1 Single SNP Analyses

The effects of each SNP on NAFLD rate were analysed individually in a series of unadjusted and unadjusted models. The results of these analyses are shown in tables 1, 2, and 3 for GoDARTS, SHARE and the meta-analysis respectively.

5.4.1.1 GoDARTS

A number of variants had significant effects on NAFLD in each cohort respectively as well as the meta-analysis. In GoDARTS, the *GLP1R* SNP rs6923761 was close to significance in the unadjusted and adjusted models, with an OR of 1.14 in each. (p = 00.058 – 0.085) In a dominant model with adjustment for age, sex and T2DM, the SNP rs1042044 had a significant association with NAFLD. (OR = 0.88(0.78 - 0.99), p = 0.037) In *GCGR*, the SNP rs28454947 was associated with NAFLD in the adjusted models, with an OR of 1.14(1.01 - 1.28) in the model adjusted for sex, age, T2DM and BMI. (p = 0.037) Full results are shown in table 19.

Table 20 - The associations between variants in GLP1R, GCG and GCGR, and NAFLD in the GoDARTS cohort

Variant	Unadjusted	Adjusted for age, sex, and T2DM	Adjusted for age, sex, T2DM and BMI	Model
rs6923761 *GLP1R*	OR = 1.14 , (0.998 - 1.3), p = 0.0538	OR = 1.14, (0.983 - 1.32), p = 0.0854	OR = 1.14 , (0.983 - 1.32), p = 0.0833	Recessive

rs1042044 *GLP1R*	OR = 0.943 , (0.849 - 1.05) , p = 0.273	**OR = 0.886 ,** **(0.788 - 0.997) ,** **p = 0.0447**	**OR = 0.88 ,** **(0.781 - 0.992) ,** **p = 0.0367**	Recessive
rs10305420 *GLP1R*	OR = 1.04 , (0.961 - 1.12), p = 0.356	OR = 1.05 , (0.965 - 1.14), p = 0.268	OR = 1.06 , (0.976 - 1.15), p = 0.166	Additive
rs140065949 *GCGR*	OR = 1.16 , (0.864 - 1.59), p = 0.329	OR = 1.18 , (0.857 - 1.65), p = 0.32	OR = 1.22 , (0.885 - 1.71), p = 0.23	Additive
rs4664447 *GCG*	OR = 1.3 , (0.937 - 1.81), p = 0.124	OR = 1.31 , (0.916 - 1.89), p = 0.147	OR = 1.34 , (0.931 - 1.94), p = 0.121	Dominant
rs28454947 *GCGR*	OR = 0.897 , (0.615 - 1.33), p = 0.579	OR = 0.734 , (0.491 - 1.11), p = 0.138	OR = 0.731 , (0.484 - 1.12), p = 0.141	Recessive
rs5386 *GCGR*	OR = 1.49, (0.731 - 3.36), p = 0.296	OR = 1.27 , (0.587 - 3), p = 0.566	OR = 1.29 , (0.592 - 3.08), p = 0.539	Recessive

5.4.1.2 SHARE

In SHARE, the *GLP1R* variant rs10305420 was associated with NAFLD in a model adjusted for sex, age and T2DM. (OR = 1.18(1.02 - 1.37), p = 0.025) The SNP rs6923761 was close to significance in the same model, with the same direction of effect as was found in GoDARTS. (OR = 1.37, (1.0 - 1.9), p = 0.059) The results of all analyses are shown in table 20.

Table 21 - Table 20 - Associations between variants in GLP1R, GCG and GCGR, and NAFLD in the SHARE cohort

Variant	Unadjusted	Adjusted for age, sex, and T2DM	Adjusted for age, sex, T2DM and BMI	Model
rs6923761 *GLP1R*	**OR = 1.38 ,** **(1.03 - 1.88),** **p = 0.0362**	OR = 1.37 , (0.996 - 1.9), p = 0.0585	OR = 1.19 , (0.749 - 1.96), p = 0.485	Recessive

rs1042044 *GLP1R*	OR = 0.973 , (0.762 - 1.23) , p = 0.824	OR = 0.892 , (0.685 - 1.15) , p = 0.39	OR = 1.01 , (0.657 - 1.5) , p = 0.977	Recessive
rs10305420 *GLP1R*	OR = 1.14 , (0.999 - 1.31), p = 0.0532	**OR = 1.18 ,** **(1.02 - 1.37),** **p = 0.025**	OR = 1.07 , (0.85 - 1.34), p = 0.586	Additive
rs140065949 *GCGR*	OR = 1.04 , (0.703 - 1.58), p = 0.857	OR = 1.14 , (0.744 - 1.79), p = 0.563	OR = 1.94 , (0.85 - 5.62), p = 0.159	Additive
rs4664447 *GCG*	OR = 0.715 , (0.479 - 1.1), p = 0.112	OR = 0.643 , (0.413 - 1.03), p = 0.0566	OR = 0.954 , (0.484 - 2.11), p = 0.899	Dominant
rs28454947 *GCGR*	OR = 0.76 , (0.51 - 1.16), p = 0.191	OR = 0.693 , (0.447 - 1.1), p = 0.108	OR = 0.823 , (0.425 - 1.76), p = 0.587	Recessive
rs5386 *GCGR*	OR = 6.14 , (1.3 - 110), p = 0.075	OR = 5.01 , (1.01 - 91.2), p = 0.12	OR = 2.09 , (0.411 - 38.2), p = 0.48	Recessive

5.4.1.3 Meta-Analysis

The IPD meta-analysis of GoDARTS and SHARE revealed a number of SNPs which were significantly associated with NAFLD. The *GLP1R* SNP rs6923761 was associated with NAFLD in the model adjusted for sex, age and T2DM. (OR = 1.15(1.01 - 1.31), p = 0.032) In the same model, rs1042044 was also associated with NAFLD. (OR = 0.88, (0.794 - 0.98), p = 0.018)

The *GCGR* variant rs140065949 was associated with increased NAFLD risk in adjusted models, with an OR of 1.40(1.04 - 1.89) in a model adjusted for sex, age, T2DM and BMI. (p = 0.029) The variant rs5386 was associated with increased NAFLD risk in an unadjusted

model. (OR = 2.06, (1.10 - 4.30), p = 0.036) This was close to significance in the model adjusted for age, sex and T2DM. All results are shown in table 4.

Table 22 - Associations between variants in GLP1R, GCG and GCGR, and NAFLD in the IPD meta-analysis of the GoDARTS and SHARE cohorts

Variant	Unadjusted	Adjusted for age, sex, and T2DM	Adjusted for age, sex, T2DM and BMI	Model
rs6923761 *GLP1R*	**OR = 1.16 , (1.03 - 1.3), p = 0.0146**	**OR = 1.15 , (1.01 - 1.31), p = 0.032**	OR = 1.14 , (0.989 - 1.31), p = 0.0722	Recessive
rs1042044 *GLP1R*	OR = 0.944 , (0.86 - 1.04) , p = 0.232	**OR = 0.881 , (0.794 - 0.978) , p = 0.0176**	**OR = 0.883 , (0.788 - 0.989) , p = 0.031**	Recessive
rs10305420 *GLP1R*	OR = 1.05 , (0.983 - 1.12), p = 0.145	OR = 1.05 , (0.984 - 1.13), p = 0.134	OR = 1.04 , (0.966 - 1.12), p = 0.291	Additive
rs140065949 *GCGR*	OR = 1.21 , (0.956 - 1.55), p = 0.12	**OR = 1.29 , (1.01 - 1.67), p = 0.0483**	**OR = 1.4 , (1.04 - 1.89), p = 0.0285**	Additive
rs4664447 *GCG*	OR = 1.04 , (0.808 - 1.35), p = 0.763	OR = 1.01 , (0.772 - 1.34), p = 0.931	OR = 1.26 , (0.915 - 1.75), p = 0.166	Dominant
rs28454947 *GCGR*	OR = 0.928 , (0.707 - 1.23), p = 0.597	OR = 0.873 , (0.654 - 1.18), p = 0.363	OR = 0.901 , (0.643 - 1.28), p = 0.55	Recessive
rs5386 *GCGR*	**OR = 2.06 , (1.1 - 4.3), p = 0.0361**	OR = 1.92 , (0.993 - 4.1), p = 0.0685	OR = 1.59 , (0.792 - 3.49), p = 0.218	Recessive

5.4.2 Analysis of Multiple SNPs

5.4.2.1 GLP1R Missense Variants

Further meta-analysis was conducted to stratify individuals into groups based on the *GLP1R*, GCG and *GCGR* variants that were informative about NAFLD risk. In the meta-analysis, the *GLP1R* missense variants rs6923761 and rs1042044 were in moderate linkage disequilibrium (LD), with a Pearson correlation of 0.62. Individuals were stratified by those who carried a NAFLD risk genotype in either/both of rs6923761 and rs1042044, and those who had a risk genotype for neither. The risk genotype for rs6923761 homozygosity for the mutant allele. (AA) The risk genotype in rs1042044 was homozygosity for the wild type allele. (AA) Individuals with at least one of these risk genotypes are henceforth referred to as being in the *GLP1R* risk group, or having the *GLP1R* risk genotype.

In GoDARTS and SHARE combined, there were 3,734 individuals with risk genotypes for either of rs6923761 or rs1042044, and 8,002 individuals who were without risk genotypes for either variants. The individuals with a *GLP1R* risk genotype for at least one of these genes had increased risk of NAFLD in the model adjusted for sex, age, T2DM and BMI. (OR = 1.15(1.04 - 1.27), p = 0.0055) The results for each model are shown in table 22, where the effect of having at least one of these risk variants versus carrying none of the risk genotypes is shown.

Table 23 - Association between the GLP1R risk genotype and increased NAFLD risk

Cohort	Unadjusted	Adjusted for age, sex, and T2DM	Adjusted for age, sex, T2DM and BMI
GoDARTS	OR = 0.997 , (0.91 - 1.09) , p = 0.947	**OR = 1.15 , (1.04 - 1.27) , p = 0.00747**	**OR = 1.16 , (1.05 - 1.29) , p = 0.00509**
SHARE	OR = 1.19 , (0.975 - 1.47) ,	**OR = 1.26 , (1.01 - 1.58) ,**	OR = 1.09 , (0.778 - 1.54) ,

	p = 0.0904	**p = 0.0389**	p = 0.622
Meta-Analysis	OR = 1.01 , (0.927 - 1.09) , p = 0.885	**OR = 1.12 , (1.02 - 1.22) , p = 0.0138**	**OR = 1.15 , (1.04 - 1.27) , p = 0.00553**

5.4.2.2 GLP1R, GCG and GCGR Variants – Combinations and Interactions

In GoDARTS, individuals who carried risk genotypes for both *GLP1R* and *GCGR* were at the highest risk of NAFLD. They had higher NAFLD risk compared to individuals with no risk variants in *GLP1R* or *GCGR*, (OR = 1.96(1.23- 3.24), p = 0.0066) and those who carried a risk variant in just one of these genes. (OR = 1.72(1.08 - 2.86), p = 0.029)

The GCGR variant rs140065949 showed different effects in individuals with and without the *GLP1R* risk genotype. This is shown in table 6.

Table 24 - The effects of GCGR variant rs140065949 on NAFLD rate in GoDARTS, stratified by the presence of the GLP1R risk genotype

GLP1R Risk Group	Unadjusted	Adjusted for age, sex, and T2DM	Adjusted for age, sex, T2DM and BMI
Non-Risk	OR = 1.1 , (0.83 - 1.46), p = 0.529	OR = 1.14 , (0.849 - 1.55), p = 0.391	OR = 1.32 , (0.934 - 1.88), p = 0.125
Risk	OR = 1.52, (0.979 - 2.47), p = 0.0735	**OR = 1.71, (1.07 - 2.83), p = 0.0304**	OR = 1.65, (0.952 - 3.05), p = 0.0887

This *GCGR* variant was associated with increased NAFLD in those who carried the *GLP1R* risk genotype, but showed no association in non-carriers.

The interactions between the *GLP1R*, GCG and *GCGR* variants were tested for any association with NAFLD in a model adjusted for age, sex, T2DM and BMI. The results are shown in table 24.

Table 25 - Association between NAFLD and the statistical interaction of the GLP1R, and GCG and GCGR variants in GoDARTS

	rs6923761	rs1042044
rs140065949	OR = 1.28, (0.701 - 2.39), p = 0.422	OR = 1.93, (0.791 - 5.54), p = 0.179
rs4664447	OR = 0.551 , (0.288 - 1.06), p = 0.0728	**OR = 0.32,** **(0.149 - 0.701),** **p = 0.00377**
rs28454947	OR = 0.977, (0.491 - 1.96), p = 0.947	**OR = 0.426,** **(0.195 - 0.947),** **p = 0.0337**
rs5386	OR = 0.316, (0.0661 - 1.39), p = 0.131	OR = 0.23, (0.0465 - 1.33), p = 0.0789

The rs1042044 AA genotype had an interaction with both rs4664447and rs28454947 that had a significant effect on NAFLD rate, with OR's of 0.32 and 0.43 respectively.

5.4.2.3 Supplementary Analysis in DMDSC Cohort

The *GLP1R* variant rs6923761 was additively associated with increased NAFLD risk in a model adjusted for age, sex, BMI and alcohol.(OR = 1.14(1.01 - 1.28), p = 0.033) The other *GLP1R* variant for which data was available (rs1042044) showed no significant association with NAFLD, nor did the *GCGR* variants rs5386 or rs140065949. No interactions between variants were found.

5.5 Discussion

The findings of the current analysis are that genetic variants in *GLP1R* and *GCGR* have associations with NAFLD rate. Two missense variants in *GLP1R* were found to be associated with NAFLD in the meta-analysis; rs6923761 and rs1042044. In *GCGR*, the intronic variant rs140065949 was associated with increased NAFLD in adjusted models, and the SNP rs5386 associated with NAFLD in an unadjusted model. *GLP1R* rs6923761 was associated with increased NAFLD risk in the DMDSC cohort also.

The wild type AA genotype homozygotes of *GLP1R* rs1042044 were at higher risk of NAFLD in our meta-analysis. The effect was also statistically significant in GoDARTS, and had a similar effect size although not significant in SHARE. A small number of studies have investigated this variant previously, with Sheikh et al. finding higher morning cortisol levels in children carrying the variant.[352] This SNP was in partial LD (0.62) with rs6923761, and this association could represent the same signal.

Homozygous carriers of the rs6923761 variant (AA genotype) were at increased NAFLD risk in the meta–analysis. Similar effect sizes, with p-values close to significance were also seen in GoDARTS and SHARE. In the DMDSC cohort, this variant was additively associated with NAFLD.

Previous literature on *GLP1R* rs6923761 has investigated a number of parameters associated with obesity, metabolism and T2DM. Sathananthan et al. found that the rs6923761 minor allele (A) decreased responsiveness to infused GLP-1.[350] They found this variant to be associated with lower beta-cell responsiveness, as individuals with at least one copy of the minor allele excreted significantly less insulin. Other studies have found reduced insulin in minor allele carriers also, as well as lower basal levels of GLP-1.[353,354] Another study found that rs6923761 minor allele homozygotes has poorer response to gliptins for HbA1c

reduction.[355] Based on these findings, rs6923761 appears to be a loss of function variant for GLP1R activity. In line with previous literature, our findings suggests that the rs6923761 variant decreases the function of the GLP-1 receptor, and thereby increases NAFLD risk.

GLP-1 has several physiological functions which are relevant to NAFLD.[356] GLP-1 promotes glucose dependent insulin secretion, and is therefore a key factor in blood glucose homeostasis. It also decreases hepatic lipogenesis and increases fatty acid oxidation in the liver, as well as insulin sensitivity.[357,358] Research into whether GLP-1 could be an effective treatment has been undertaken. Infused GLP-1 has been found to prevent NAFLD in mice.[359]

Despite these effects of GLP-1, the administration of GLP-1 may not be an effective therapy for NAFLD. GLP-1 has a half-life of 90 seconds in the body, making it unviable as a drug in that form.[360] Further to this, NAFLD patients have been found to have normal GLP-1 levels, despite decreased incretin effect.[361] This may be due to the GLP-1 resistance which has been found in NAFLD.[362] GLP-1 receptors have been found to be downregulated in individuals with NAFLD, causing this resistance and lowering the effect of GLP-1.[363] DPP4, which deactivates the GLP-1 enzyme, has also been observed to be higher in patients with NAFLD, further decreasing the effectiveness of GLP-1.[364]

To achieve the desired increase in incretin effect, agonism of the GLP-1 receptor has been targeted with a number of drugs.[365] The increased effect of GLP-1 associated with GLP1R agonists increases insulin sensitivity and production.[102] GLP1R agonists have shown positive effects on NAFLD in a number of studies, with a study in mice showed that GLP1R agonists reduced oxidative stress as well as hepatic fat.[359,366] The improvement in glycaemia seen with the administration of GLP1R agonists correlated with reduction in liver fat observed.[366] Further to this, GLP1R agonism exhibited anti-inflammatory effects in the liver.[367] This is via the downregulation pro-inflammatory cytokines and transcription factors.[368] Decreasing

hepatic steatosis and inflammation are key factors in the treatment of NAFLD and averting serious liver damage associated with progressive stages.

The current study found that patients with loss of function variants in *GLP1R* had increased NAFLD risk, which is consistent with previous research suggesting agonism of GLP-1R should be beneficial for NAFLD. Patients with different *GLP1R* genotypes may have different response to GLP1RAs. These findings may be used for applied personalised medicine in those with NAFLD, matching patients to the drugs and doses which would be the most effective and safe.

Significant associations between NAFLD and genetic variants in GCGR; the gene which encodes the receptor for glucagon; were found. We found the *GCGR* variant rs140065949 was associated with increased NAFLD risk in the meta-analysis, in a model adjusted for sex, age, T2DM and BMI. This SNP is intronic and doesn't appear in any previously published studies. This is also true of the synonymous variant rs5386 which associated with increased NAFLD in an unadjusted model. In a recessive model, rs5386 was close to significance in the model adjusted for sex, age, and T2DM in the meta-analysis, and also close to significance in the unadjusted analysis in SHARE. The intronic GCG variant rs4664447 had a significant interaction with rs1042044 in *GLP1R.* In carriers of the rs1042044 C allele (non-NAFLD risk allele), rs4664447 was associated with increased NAFLD risk. However, in those who were homozygous for the wild type rs1042044 A allele, rs4664447 was associated with an almost 50% reduction in NAFLD risk. A previous study found variation in rs4664447 was associated with decreased insulin, GLP1R and glucagon levels, although this was for the T>G variation unlike the T>C variation analysed in the current study.

Glucagon is a hormone which primarily acts to raise glucose and fatty acid concentrations in the bloodstream.[346] It has a number of significant effects in the liver. Glucagon stimulates the

glycogenosis in the liver, where stored glycogen is released into the bloodstream in the form of glucose.[369] It also stimulates gluconeogenesis and lipolysis in the liver, as well as decreasing rate of fatty acid synthesis.

The findings of previous research combined with the current study suggest a role of the *GCGR* gene in NAFLD, and warrant further research into the functions of these variants and their relationship with NAFLD. NAFLD has been linked to GCGR signalling. Kazda et al. found that GCGR antagonism increases hepatic steatosis.[370] Nason et al. found that reduced GCGR signalling increases fatty acid oxidation in the liver, and reduced liver triglyceride levels.[371] It has also been shown that a high fat diet reduces glucagon receptor content in rat livers. Glucagon resistance has been suggested as a cause of lowered effect of GCG in the liver for glucose production. Despite several effects which are associated with poor glycaemic control, including high blood glucose and hyperinsulinemia, agonism of the glucagon receptor has been studied as a therapy for various metabolic conditions.[372,373] These include thermogenic and catabolic effects.[371] GCGR is associated with increased energy expenditure and reduced food intake.[374] This effect has been combined with GLP1R agonism, in a number of GLP1R/GCGR co-agonist drugs.[375]

GLP1R and GCGR co-agonists are a class of drug which has been investigated as a therapy for obesity and T2DM. Farooq et al. showed enhanced insulin as a results of GLP1R/GCGR co-agonism, and obesity reducing effects have been shown in rodents.[102,103] They can also lower cholesterol and increase insulin sensitivity.[347] Elvert et al. found mixed results in crab-eating macaque monkeys, as GLP1R/GCGR co-agonism increased weight loss and insulin secretion, but was associated with lower glycaemic control long term.[376] These effects make GLP1R/GCGR co-agonists worthy of further investigation for treatment of number of obesity related conditions.

Co-agonism for GLP1R and GCGR has also been investigated as a therapy for NAFLD, and shown promising results. Patel et al. found that co-agonism for these receptors reduced hepatic steatosis as well as lipotoxicity.[97] Similar findings have also been made in mice.[375] As well as simple steatosis, GLP1R/GCGR co-agonists have been found to improve NASH and fibrosis.[101] Kannt et al. found that a GLP1R/GCGR co-agonist reduced the histological severity of NASH.[377] As well as their individual effects, GLP-1 and glucagon are linked as GLP-1 inhibits glucagon secretion when blood glucose levels are raised.[361]

The results of our study are consistent with this, as we show that loss of function of *GLP1R* increases risk of NAFLD, and therefore it can be extrapolated that upregulation of GLP1R will reduce NAFLD. We also showed effects of variants in *GCGR*, which further advocates a beneficial effect of GLP1R and GCGR co-agonism on NAFLD. We found that individuals who carried risk variants in both *GLP1R* and *GCGR* were at greater risk than those who carried no risk variants, or just one risk variant. The *GCGR* variant rs140065949 had a significant effect on NAFLD risk only in those who carried the *GLP1R* risk genotype. The *GLP1R* variant rs1042044 also had a significant interaction with two *GCGR* variants, which suggests interplay between the pathways associated with these two genes.

Stratification by genotype may be an effective means of ensuring optimal treatment, should GLP1R/GCGR co-agonists become widely used. Individuals with different *GLP1R* and *GCGR* genotypes may benefit differently or not at all from GLP1R/GCGR co-agonists. With the advent of precision medicine, this may help ensure individuals are treated as effectively and economically as possible.

The main goal of this study was to analyse the effects of *GLP1R*, *GCG* and *GCGR* variants on NAFLD in two Scottish cohorts. Analysis of these variants in the DMDSC cohort did however show that the *GLP1R* rs6923761 variant has effects on NAFLD in South Asians

also. Previous research has shown improvements a number of clinical outcomes including HbA1c and ALT levels in an Indian population with a GLP-1R agonist.[378] This combined with research showing effects of the NAFLD risk variant *PNPLA3* rs738409 in European and South Indian, as well as a previous chapter in the current thesis showing the same, suggest at least some commonality in the genetic risk architecture and disease pathology between Caucasian and South Indian individuals.[256]

5.5.1 Limitations

A limitation of the present study is the NAFLD phenotype. This was defined by the presence of raised ALT levels. There is evidence of NAFLD in individuals with normal ALT levels in some studies, especially in South Asians.[379] BMI is also less effective as a predictor of NAFLD in South Asian populations.[380] However, overall ALT is an effective and practical means for defining NAFLD in large populations.[381] The longitudinal nature of the data used in this study improves the accuracy of the definition also, as does using the ULN's suggested by Prati et al., which are lower than many previously used limits, as this increases the sensitivity of the definition.[180] The missingness of BMI data for individuals without T2DM in SHARE also reduced the cohort size available for analysis, and lowered experimental power.

The replication of results in the DMDSC cohort has some value, as it shows that the *GLP1R* rs6823761 variant has a significant effect on NAFLD. The power of this analysis is low however, with only 3,154 participants, and data for only four out of the seven variants was available.

The difference in NAFLD phenotype ascertainment between DMDSC and the GoDARTS and SHARE cohorts limit the comparability of results. This means that it is not possible to determine whether the lack of association between the *GLP1R* rs1042044 variant and NAFLD in DMDSC is due to true differences in the populations, or due to these confounding

factors. Compared with GoDARTS and SHARE, the NAFLD phenotype was more likely to include patients with non-NAFLD liver insults as cases. This was due to lack of availability of data for this. Alcohol was controlled for in this analysis as a binary trait, but there was no indication in the data whether individuals consumed harmful amounts of alcohol, likely to cause liver disease. Likewise, the data did not have sufficient information to be able to exclude individuals with alternate causes of liver disease such as virological or immunological insults.

5.6 Conclusion

We found effects of a number of genetic variants in *GLP1R* and *GCGR* on NAFLD. Two missense variants in *GLP1R* were associated with NAFLD rate. There is evidence in previous literature that these variants decrease the activity of GLP-1 receptors, supporting the notion that GLP-1 receptor agonism may be beneficial for NAFLD. The *GCGR* variant rs140065949 was associated with increased NAFLD risk in the adjusted models. This likewise suggests a role of GCGR in NAFLD. Individuals who had risk genotypes in both *GLP1R* and *GCGR* were at greatest NAFLD risk, and the effect of the *GCGR* variant was only seen in carriers of the *GLP1R* risk genotype. These results combined with previous studies suggest that co-agonism of GLP1R and GCGR may be an effective therapeutic approach for NAFLD. Stratifying individuals by the genetic variants we found to affect NAFLD may aid treatment of NAFLD and other obesity related conditions.

6 Endothelin Genes and NAFLD

6.1 Abstract

Endothelin is a vasoconstrictor which has significant effects in the liver, including modulating hepatic glucose output, and increasing risk of non-alcoholic steatohepatitis (NASH). Endothelin receptor antagonists (ETRA) have been studied in animal models as a potential NASH treatment. The aim of this study was to investigate the role of genetic variants which affect endothelin and endothelin receptor expression in NASH.

Analyses for this cohort study were conducted in the GoDARTS cohort, a Scottish cohort of 13,695, individuals. Replication was conducted in the SHARE cohort, with 3,068 individuals, and in the MDRF cohort with 3,068 genotyped individuals. The primary outcome of the study was Non-Alcoholic Steatohepatitis (NASH), which was defined by diagnoses in medical records. Genetic variants in *EDN1, PHACTR1, EDNRA* and *EDNRB* were selected for analysis based on previous literature linking them to endothelin and endothelin receptor expression respectively.

In GoDARTS, three variants in *EDNRA* were significantly associated with NASH; rs17612742 – OR = 1.59(0.99 -4.02), p = 0.04); rs1878406 and rs6841581– OR = 1.81(1.16 – 2.73), p = 0.006). In SHARE, FIB4 index was significantly associated with the same variants in *EDNRA*; rs6841581 – β= 0.25(0.077 -0.42), $p = 4.6x10^{-3}$; rs1878406 – β = 0.26(0.087 – 0.43), $p = 3.2x10^{-3}$; rs17612742 – β = 0.24(0.075 – 0.41), $p = 4.7x10^{-3}$. A number of other associations between endothelin SNPs and NASH related phenotypes were found.

Genetic variants which are known to affect endothelin and endothelin receptor expression have significant effects on NASH and related phenotypes, including portal hypertension. These findings have relevance to research into ETRAs as NASH treatment, and to research

into the understanding of the pathogenesis of NASH in general. They may be useful for genetic stratification with respect to therapeutic intervention in NASH.

6.2 Introduction

Non-alcoholic fatty liver disease (NAFLD) is the most common cause of liver disease globally. It is estimated to affect around 25.2% of adults worldwide.[4,34] It is a prevalent comorbidity of obesity, and frequently occurs in individuals type 2 diabetes (T2DM).[382] There is currently no specific pharmacological intervention recommended for NAFLD, though a number of drug targets and pathways are currently being investigated, including antagonism of endothelin-1. (ET1)

Endothelin is a vasoconstrictor, which when overexpressed contributes to hypertension, heart disease, and a number of other conditions.[383,384] Endothelin acts on endothelin receptors, which come in two main types; Endothelin receptor A (ETA) and Endothelin receptor B (ETB).[385] When activated ETA receptors' main role is vasoconstriction. ETB receptors on the other hand have a role in vasodilation, as nitric oxide is released when they are activated.[386]

Several genes are associated with the production, reception and action of endothelin. *PHACTR1* has been found to regulate endothelin expression, by regulating the endothelin 1 gene.[114] (*EDN1)* The *EDN1* gene itself has several variants which influence levels of endothelin in the body, and have been associated with clinical outcomes.[112] The endothelin receptor A and B genes (*EDNRA* and *EDNRB*) each encode their respective receptor.[113]

Endothelin has significant effects in the liver, and is a key regulator of hepatic blood flow. The liver has an important role in the removal of ET1 from the bloodstream and in patients with liver disease, including NASH, high serum levels of ET1 have been observed.[384,387] As well as alterations in hepatic blood flow, ET1 is associated with increased activation and

proliferation of hepatic stellate cells (HSCs). This is important in NAFLD as HSCs are the main producers of the extracellular matrix that characterises hepatic fibrosis.[111,388]Increased glycogenolysis and hepatic glucose output are also associated with ET1.[387] In rats and in humans with cirrhotic livers, increased expression of endothelin receptors has been found.[109,389] These findings demonstrate an association between the ET1 pathway in NASH and fibrosis.

Following on from research into the link between endothelin and NASH, investigations into the use of endothelin receptor antagonists (ETRA) have been conducted. . The main role of these drugs is the reduction of pulmonary hypertension, as they block the vasoconstrictive effect of ET1. [108,390] ETRAs have been shown to reduce liver fibrosis in rats.[391] Similar results were found in mice, as the ETRA ambrisentan reduced progression of hepatic fibrosis by inhibiting hepatic stellate cell activation.[107] Some promising effects of ETRAs for treating NASH in a small human study have been found, but further research is required.[392] Figure 6-1 below illustrates the interaction between endothelin, its receptors and the processes of vasoconstriction and hepatic stellate cell activation, as well as the site of ETRA action.

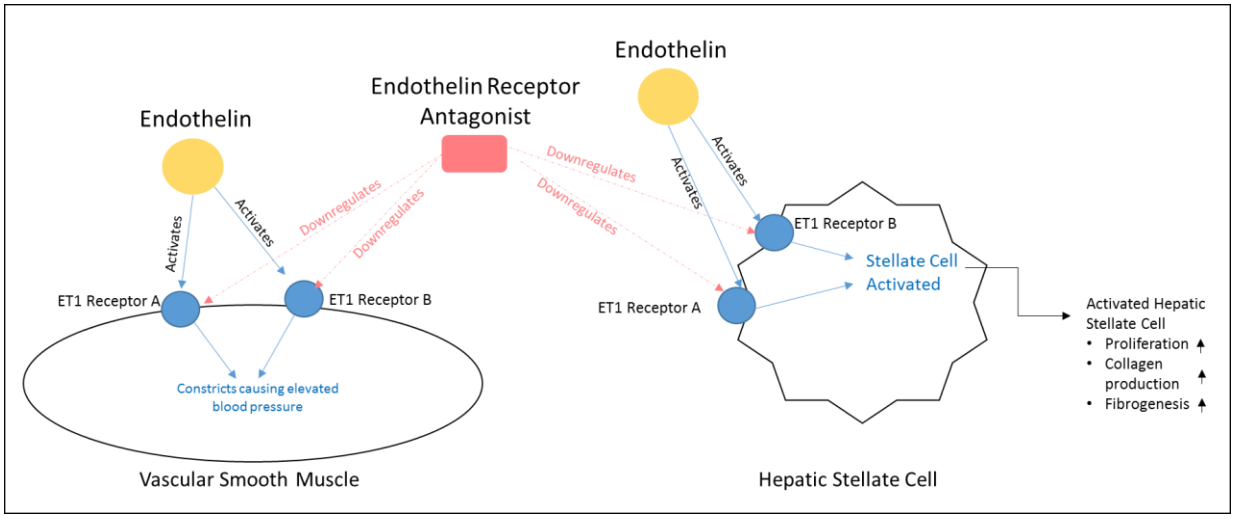

Figure 6-1Endothelin Action on Vascular Muscle and Hepatic Stellate Cells

The aim of this study was to investigate the relationship between endothelin related genetic variants and NASH related outcomes. Genetic variants known to correlate with endothelin and endothelin receptor (ETR) activity were selected for this analysis. The main outcome phenotype was NASH, fibrosis and NAFLD hospitalisation also included. A significant effect of ET1 or ETR related variants would further clarify the role of the endothelin pathway in NASH, and may be useful for stratification of patients if an ETRA therapy is approved. To investigate whether any associations seen were due to haemodynamic effects or direct action on hepatic stellate cells, portal hypertension and alcoholic liver disease were also analysed.

The genetic variants to be analysed were selected from previous literature linking genetic variants to ET1 and ETR expression. Three common variants from *ET1* were analysed; rs1800541, rs2070699, and rs5370. These variants have been linked to subarachnoid haemorrhage, hypertension and ischemic stroke.[113,393] Also analysed was rs9349379 variant in *PHACTR1*, which has been linked to ischemic heart disease. [114] This variant is close to the *EDN1* locus, and is known to regulate ET1 expression.[394]

Variants associated with endothelin receptors were analysed. Variants in *EDNRA* were investigated as this gene has been linked to conditions including hypertension.[112,395] The SNPs rs6841581, rs17612742, and rs1878406, which are in very high linkage disequilibrium, have shown association with ischemic stroke and ischemic heart disease, through increased ETR activity.[115,300,396] The SNP *EDNRA* rs4593108 has also shown associations with ischemic heart disease, through ETR activity increases.[397] Common variants in *EDNRB*, which has been associated with Hirschsprung disease, were also selected for analysis; rs3818416 and rs5351.[398]

6.3 Methods

6.3.1 Data

The GoDARTS study was used as a discovery cohort.[187] This is a T2DM case-control cohort with electronic health record data available for the 18,306 participants and genetic data available for many of these. The SHARE cohort was also analysed, and had 3,068 individuals with sufficient phenotypic and genotypic data available for analysis. The DMDSC cohort was included in further analysis, and had 2,013 individuals with adequate data for analysis. Full descriptions of these cohorts and phenotypic outcomes are presented in a previous chapter of the current thesis.

6.3.2 SNPs

Candidate SNPs which have previously been linked to ET1 and ET_A activity were selected for analysis. These are shown in table 25 below, with frequencies from dbSNP.[399]

Table 26 - Endothelin and Endothelin Receptor related genes; numbers and Minor Allele Frequencies (MAF)

Gene	SNP	GoDARTS Allele Frequencies			GoDARTS MAF	SHARE MAF	MDRF MAF	dBSNP Reported
		0	1	2				
EDN1	rs1800541	4424	1888	224	0.179	0.177	0.243	0.246
EDN1	rs5370	5838	3555	579	0.236	0.232	0.367	0.228
EDN1	rs2070699	1823	3016	1270	0.455	0.474	0.363	0.443
PHACTR1	rs9349379	3414	4606	1675	0.41	0.409	0.490	0.401
EDNRA	rs4593108	4494	1909	221	0.177	0.174	0.363	0.176
EDNRA	rs17612742	7277	2306	209	0.139	0.137		0.137

							0.206	
EDNRA	rs1878406	7258	2344	213	0.141	0.14	0.203	0.146
EDNRA	rs6841581	7267	2334	217	0.141	0.141	0.221	0.148
EDNRB	rs3818416	374	2454	3783	0.756	0.753	0.805	0.748
EDNRB	rs5351	988	3187	2442	0.610	0.605	0.583	0.574

A number of these variants were in high linkage disequilibrium (LD), which is shown in the correlation matrix in table 26. A number of other *EDNRA* SNPs were in extremely high LD with the SNPs of interest ($r^2 > 0.995$), therefore were not analysed. These included: rs6842241, rs1801708, rs11413744, rs6537481, rs786205230, and rs6841473.

Table 27 - Correlation matrix of SNPs in EDNRA in the GoDARTS cohort

	rs4593108	rs17612742	rs1878406	rs6841581
rs4593108	1			
rs17612742	0.0662	1		
rs1878406	0.0646	0.964	1	
rs6841581	0.0676	0.968	0.994	1

The two SNPs in *EDRNB* which were analysed were in moderate LD, with a Pearson's R^2 of 0.70.

Further to these five SNPs, three variants known to associate with NAFLD and NASH risk were analysed in the model with NASH in GoDARTS to test for any interactions or

alterations of associations with NASH and endothelin related SNPs. These were *PNPLA3* rs738409, *TM6SF2* rs58542926 and *HSD17B13* rs6834314, shown in table 27.

Table 28 - NAFLD and NASH related genes from previous literature; numbers and MAFs.

Gene	SNP	0	1	2	GoDARTS MAF	dBSNP Reported
PNPLA3	rs738409	6797	3417	446	0.202	0.278
TM6SF2	rs58542926	11220	1664	76	0.0701	0.0653
HSD17B13	rs6834314	4673	3557	690	0.277	0.251

6.3.3 Outcomes

6.3.3.1 NASH

The main phenotype investigated in this study was NASH. This was defined by the presence of the relevant NASH ICD-10 codes in admissions and deaths records for participants in GoDARTS at any point in their life. These were: "K75.8", "K740.", "K74.1", "K74.2", "K72.9", and "K74.6".[225] Similar methods of using an ensemble of ICD-10 codes to classify NAFLD and its subsequent stages have been used previously, using codes for NAFLD, NASH, fibrosis, cirrhosis, cryptogenic cirrhosis and unspecified hepatic failure.[59,400,401]

6.3.3.2 Fibrosis

Fibrosis was also investigated in this study. This was defined by the presence of the relevant fibrosis ICD-10 codes in admissions and deaths records for participants in GoDARTS at any point in their life. These were the same as the NASH codes, with the omission of the “K75.8” code.

6.3.3.3 NAFLD Hospitalisation

This was defined by the presence of NAFLD ICD-10 codes, or subsequent stages of NAFLD including NASH and fibrosis in admissions and deaths records ever. The ICD-10 codes used were “K76.0”, plus the codes used for NASH: "K75.8", "K74.0", "K74.1", "K74.2", "K72.9", and "K74.6".

6.3.3.4 Ischaemic Heart Disease

As a positive control to validate the effects of the genetic variants of interest, associations with ischaemic heart disease were tested. This condition was phenotyped using admissions and deaths data, which were searched for the occurrence of any ICD-10 code relating to ischemic heart disease. These were: “I20”, “I21”, “I22”, “I23”, “I24”, “I25”, and “I26”. In GoDARTS, 2,789 individuals had a diagnosis of ischemic heart disease at some point.

6.3.3.5 Portal Hypertension

To investigate the mechanism by which endothelin and endothelin receptor variants potentially altered NASH risk, portal hypertension was analysed. This was phenotyped with the ICD10 code “K76.6”.

6.3.3.6 FIB-4 Index

The FIB-4 index was developed as a non-invasive assessor of fibrosis, and FIB-4 score over 3.25 is known to correlate with high likelihood of advanced fibrosis.[153] This was calculated using individuals’ most recent measurements for the component biomarkers. When analysed as a continuous variable, FIB-4 was log transformed, as it was not normally distributed.

6.3.3.7 ALT to AST Ratio

ALT to AST ratio was calculated using patients most recent ALT and AST measurements prior to sign up to GoDARTS. ALT to AST ratio is a commonly used biomarker for liver

damage, particularly NAFLD.[402,403] ALT and AST measurements were analysed versus the selected genetic variants separately, as well as a ratio.

6.3.3.8 APRI

The APRI score is an index of fibrosis which is calculated as the ratio of AST to platelet count.[152] It has been shown to accurately diagnose hepatic fibrosis.

6.3.4 Outcomes in MDRF Cohort:

6.3.4.1 NAFLD

NAFLD defined by the presence of elevated ALT levels was one of the phenotypes present in the MDRF cohort data. Patients in the MDRF cohort had a number of biochemical markers measured on their first visit to the clinic, and ALT was measured at this point.[180] The thresholds for ALT levels in this study were 19U/L for women, and 30U/L for men.[180]

6.3.4.2 ALT to AST Ratio

Similarly to the previous ALT based NAFLD definition, this measure uses biochemical measurements taken at patients' first visit to the MDRF clinic. ALT and AST measurements were analysed versus the selected genetic variants separately, as well as a ratio.

6.3.4.3 Fatty Liver Grade

Fatty Liver Grade (FLG) was assessed by abdominal ultrasound, and is a measure of the fatty infiltration of the liver.[196] FLG was measured on a scale of 0 to 3, with these levels representing degrees of fatty infiltration of the liver; none, mild, moderate and severe respectively. This was analysed as a continuous trait, and also dichotomised as 0 versus 1, 2 and 3; i.e. no fatty infiltration of the liver versus any level of fatty infiltration of the liver.

6.3.5 Exclusions

To ensure the specificity of the NAFLD definition individuals with features of alternate causes of liver disease were excluded. Individuals with any positive serological tests for anti-smooth muscle antibody, antinuclear antibodies or anti-mitochondrial antibodies, any positive serology for hepatitis B surface antigen or hepatitis C antibody, or mention of cause of liver disease in medical records were excluded. Individuals with alcohol dependence or any documentation of alcoholic liver disease in their EHRs were excluded. In addition, individuals who self-reported drinking more than 20g a day for women and more than 30g a day for men were excluded.

6.3.6 Statistical *Methods*

All analyses were adjusted for age, sex, and type 2 diabetes (T2DM). Continuous variables were analysed using a linear regression and binary variables were analysed with a logistic regression. All analyses were carried out in the statistical package R. This was a cross sectional study, as outcome phenotypes were assessed as the presence of a condition ever before the last follow up date.

6.4 Results

6.4.1 Results in GoDARTS

Genetic variants associated with ET1 and ET_A were tested for associations with NASH, and related phenotypes. A number of endothelin related SNPs in or near *EDNRA* and EDN1 had significant associations with NASH, and NASH related phenotypes. The key findings from this study are reported below.

6.4.1.1 Ischemic Heart Disease

Positive control tests were run with genetic variants associated with ET1 and ET_A, and coronary heart disease. The *EDNRA* variant rs6841581 showed an association with ischaemic heart disease, with an odds ratio of 1.11(95% CI = 1.00 - 1.23), p = 0.039), as did rs1878406. (OR = 1.11(1.01 - 1.23), p = 0.037)

6.4.1.2 NASH

Three variants in *EDNRA* were significantly associated with NASH; rs17612742 – OR = 1.59(0.99 -4.02), p = 0.04); rs1878406 and rs6841581 (due to high LD) – OR = 1.81(1.16 – 2.73), p = 0.006);

In a recessive model, *EDNRA* rs4593108 *was* non-significantly associated with NASH, with similar OR to other *EDNRA* variants. (OR =1.86(0.65 - 4.22), p = 0.18)

The known NAFLD and associated condition risk variants *PNPLA3* rs738409, *TM6SF2* rs58542926 and *HSD17B13* rs6834314 were added to the models to test for any interactions with *EDNRA* variants. It was found that the *EDNRA* variants which associated with NASH behaved additively with these variants, and no interaction was found.

6.4.1.3 Fibrosis

None of the variants analysed were significantly associated with fibrosis.

6.4.1.4 NAFLD Hospitalisation

EDNRA variants were significantly associated with NAFLD hospitalisation. These were: rs1878406 – OR = 1.46(1.04 – 2.00) p = 0.023; rs6841581 – OR = 1.46(1.04 – 2.00), p = 0.024. *EDNRA* rs17612742 and rs4593108 were both close to statistical significance for NAFLD hospitalisation, with similar odds ratios as seen for NASH.

6.4.1.5 Fatty Liver Index

The variant *EDNRA* rs6841581 was associated with increased FLI: rs6841581 – β = 1.66(SE = 0.87), p = 0.048. The other two variants in *EDNRA* which are in high LD with rs6841581 (rs17612742 and rs1878406) were close to statistical significance with similar beta estimates.

6.4.1.6 GGT

Three variants in *EDNRA* were associated with increased serum GGT: rs17612742 - β = 10.04(SE = 3.00), p = 0.00084; rs1878406 - β = 10.35(SE = 3.02), p = 0.00062; rs6841581 – β = 10.76(SE = 3.01), p = 0.00036)

6.4.1.7 FIB4

EDNRA rs4593108 was associated with increased FIB4 index in a recessive model. (β =0.79(SE = 0.37), p = 0.031.

EDN1 rs2070699 was associated with lower FIB4 index. (β = -0.77(SE = 0.31), p = 0.012)

6.4.1.8 APRI Score

EDNRA rs4593108 was recessively associated with APRI score. (β = 0.34(SE = 0.12), p = 0.0048)

6.4.1.9 Portal Hypertension

Variants in *EDNRA* were associated with increased portal hypertension. rs6841581 – OR = 2.40(1.014 -5.16), p = 0.032 and rs1878406 – OR = 2.42(1.017 - 5.21), p = 0.031) *EDNRA* rs4593108 was close to statistical significance for association with portal hypertension. (OR = 2.04(0.85 - 4.48), p =0.089)

6.4.1.10 ALT to AST Ratio

None of the variants analysed were associated with ALT to AST ratio.

6.4.2 Analysis in SHARE Cohort

To validate findings, a number of phenotypes were analysed in SHARE, with the *EDN1, EDNRA*, *EDNRB* and *PHACTR1* SNPs of interest.

6.4.2.1 NASH

Homozygous carriers of *EDNRA* rs4593108 had increased risk of NASH. (OR = 4.04(0.92 - 1.24), p = 0.029)

PHACTR1 rs9349379 was recessively associated with increased NASH risk. (OR = 2.39(1.02 - 5.24), p = 0.034)

6.4.2.2 Fibrosis

Homozygous carriers of *EDNRA* rs4593108 also had increased risk of fibrosis.(OR = 4.54(0.70 - 6.94), p =0.05)

PHACTR1 rs9349379 was recessively associated with increased fibrosis risk. (OR = 2.41(0.97- 5.53), p = 0.044)

6.4.2.3 NAFLD Hospitalisation

EDNRA rs4593108 was associated with increased NAFLD hospitalisation. (OR = 4.05(0.95 - 12.0), p = 0.026)

For the variants in *EDNRA* that were significant in GoDARTS, although the p values were not significant, the direction of effect was the same as in GoDARTS. For example for NAFLD Hospitalisation: rs1878406- OR = 1.36(0.84 - 2.1297), p = 0.19; rs6841581 – OR = 1.36, (0.83 - 2.12), p = 0.20; rs17612742 - OR = 1.39(0.85 - 2.17), p = 0.16.

6.4.2.4 FIB4

FIB4 index was significantly associated with variants in *EDNRA*, with the results as follows. rs6841581 – β= 0.25(0.077 -0.42), p = 4.6×10^{-3}; rs1878406 – β = 0.26(0.087 – 0.43), p =

$3.2x10^{-3}$; rs17612742 – β = 0.24(0.075 – 0.41), p = $4.7x10^{-3}$. *PHACTR1* rs9349379 was recessively associated with increased FIB index. (β =4.22(SE = 2.12), p = 0.47)

These variants were also associated with greater odds of having experienced FIB4 index greater than 3.25. The results were as follows: rs6841581 – OR = 2.24(1.20 - 4.10), p = $9.2x10^{-3}$; rs1878406 – OR = 2.30(1.24 - 4.19), p = $7.2x10^{-3}$; rs17612742 – OR = 2.32(1.27 - 4.16), p = $5.0x10^{-3}$.

6.4.2.5 Portal Hypertension

Variants in *EDNRA* were associated with increased portal hypertension. rs6841581 – OR = 2.39(1.31 - 4.17), p = 0.003; rs17612742 – OR = 2.44(1.34 - 4.25), p = 0.0023; rs1878406 – OR = 2.40(1.31 - 4.18), p = 0.0029.

6.4.3 Analysis in MDRF Cohort

6.4.3.1 NAFLD

A single variant in *EDNRA* (rs4593108) was close to statistical significance for NAFLD. (OR = 0.90134 (0.80 -1.01), p = 0.085)

6.4.3.2 ALT to AST Ratio

None of the variants tested were significantly associated with ALT to AST ratio.

Two variants in *EDN1* were associated with ALT level in recessive models; rs5370 –β= 3.66(1.36 - 2.69), p = 0.0073 and rs2070699 – β= -2.70(1.29 - 2.09), p = 0.037

6.4.3.3 Fatty Liver Grade

Two variants in *EDNRA* were close to statistical significance for a Fatty Liver Grade Greater than 0. (rs17612742 - OR = 1.24(0.99 - 1.58), p = 0.07; rs1878406 – OR = 1.22(0.97 - 1.54), p = 0.088;

6.5 Discussion

We found significant associations between genetic variants in endothelin and endothelin receptor A, and NASH, plus NASH related phenotypes in GoDARTS and SHARE. In the MDRF cohort, we found variants significantly associated with ALT levels, and close to significance for NAFLD and FLG. A number of these variants have been linked to upregulation of the endothelin pathway in previous studies.

We found two independent variants in *EDNRA* to be associated with a number of the NASH related outcomes that were analysed in this study. The first locus contained rs17612742, rs1878406 and rs6841581, all of which were in high LD. (Pearson's $R^2 = 0.96$ -0.99) Associations were found for these SNPs for NASH, NAFLD hospitalisation, FLI, portal hypertension and ischemic heart disease in GoDARTS. In SHARE, associations were found for FIB4 index and portal hypertension. Although the same direction of effect for these SNPs was seen in GoDARTS and SHARE for associations with NASH, fibrosis and NAFLD hospitalisation, none were statistically significant, likely due to low power. In the MDRF cohort, rs17612742 and rs1878406 were close to significance for association with increased risk of Fatty Liver Grade greater than 0. i.e presence of fatty infiltration of the liver.

This locus is known to associate with both carotid intima media thickness (cIMT) and plaque, and is thought to upregulate the activity of the endothelin receptor.[404] GWAS studies have found that this locus is associated with increased systolic blood pressure, and increased pulse pressure.[312] It has also been associated with increased risk of coronary artery disease, thought to be a result of increased cIMT and plaque.[405] Association between this *EDNRA* locus and GGT was observed. GGT is often used as a marker of liver damage, but is also associated with ischemic heart disease.[148,406] In SHARE, these three *EDNRA* SNPs were significantly associated with increased FIB-4 index, a useful marker of fibrosis. These findings are

consistent with this *EDNRA* locus being associated with increased endothelin receptor activity.

The second locus was at *EDNRA* rs4593108, which was in extremely low LD with the previous *EDNRA* locus around rs17612742. (Pearson's $R^2 < 0.07$) In GoDARTS, we found associations with this SNP for FIB4 and APRI. It was also close to significance for NASH, NAFLD hospitalisation and portal hypertension. In SHARE, it was associated with NASH, NAFLD hospitalisation, and fibrosis. In MDRF, it was also close to significant association with increased NAFLD. (As defined by raised ALT levels) This locus has been reported to associate with a number of clinical outcomes, including coronary heart disease.[397] Studies have also found that it is associated with increased risk of myocardial infarction.[407] These findings suggest that this variant is associated with upregulation of endothelin A receptors.

Significant associations between the *PHACTR1* variant rs9349379 and NASH, fibrosis and FIB index were found in SHARE, but not replicated in GoDARTS or MDRF. This variant has been found to associate with increased coronary artery disease risk in a number of studies.[408] It is also associated with carotid plaque. [409] It has recently been demonstrated that *PHACTR1* rs9349379 regulates the *EDN1* gene.[114] The minor allele carriers were found to have higher *EDN1* expression, and higher ET1 levels.

The findings of the current study suggest that the upregulation of endothelin and the endothelin A receptors plays a major role in development of NASH and fibrosis. Leivas et al. found that expression of ETA and ETB receptors was increased in the livers of cirrhotic patients, and was directly associated with portal hypertension in these individuals. Tsuchiya et al. found that in cirrhotic rats, higher endothelin activity was associated with hepatic ischemia and reperfusion injury, as well as lowered survival.[410] Serum levels of endothelin are increased in cirrhotic patients, and the levels of endothelin correlate with the severity of

liver damage as well as portal hypertension.[411] These findings, combined with our results linking endothelin and endothelin receptor related genetic variants suggest an important role for endothelin in liver disease, especially fibrosis.

Endothelin is associated with a number of effects and pathways which may be responsible for these associations, and may contribute to the pathophysiologies of NASH and fibrosis.

Endothelin is a vasoconstrictor, and acts to increase blood pressure.[383] *EDNRA* (Endothelin receptor A) is a sub-type of endothelin receptor which increases vasoconstriction when activated, which raises blood pressure.[412] Haemodynamic changes and effects are present in fibrosis and cirrhosis.[413] Hypertension of the hepatic artery is caused by increased extrahepatic hyperdilation, and the resulting increased inflow of blood.[414] This increase in vasodilators which cause this hyperdilation is thought to be due to portosystemic shunting and bacteria translocation.[415] Portal hypertension is a key element of cirrhosis. Additionally to increased inflow to the portal vein, portal hypertension is increased by the build-up of extracellular matrix causing increased resistance to blood flow, as well as active vasoconstriction within the liver.[416] Portal hypertension is a major factor in the development of complications in liver disease.[417] Moller et al. found increased circulating endothelin in patients with cirrhosis, causing significant changes in hepatic haemodynamics.[411,415] Endothelin has been shown to decrease splanchnic blood flow.[418] There a number of pathways which endothelin could affect to alter splanchnic circulation including through nitric oxide(NO), arachidonic acid metabolites or bacteria.[415,419,420]

Hepatic ischemia is a feature of severe liver disease, and in particular microvascular circulation is impaired in NAFLD.[421] It has been found that steatotic livers are less tolerant of hepatic ischemic reperfusion injuries.[422] Tsuchiya et al. found increased the ischemia of the

liver seen in cirrhotic rats was reduced with the administration of ETA receptor antagonists.[410,423]

Endothelin is associated with increased secretion of inflammatory cytokines.[424] These include TNF-α, IL-1 and IL-6, which are important inflammatory markers in the development of NASH.[425] As well as hepatic inflammation, these cytokines have a role in increasing the extracellular matrix production associated with fibrosis.[424] The ET1/ETA pathway is associated with increased myocardial fibrosis, through fibroblast proliferation and extracellular matrix deposition.[426] Increased ET1 levels, as well as increased numbers of ETA receptors have also been seen in patients with pulmonary fibrosis.[427] The increased heart and lung fibrosis associated with ET1 may have mechanisms in common with the increased NASH and fibrosis observed in the current study.

Hepatic stellate cells (HSC) play a major role in fibrosis, and are an important factor in the association between endothelin and fibrosis.[428] HCSs are normally inactive and have a role in the storage of vitamin A. When activated in response to liver damage, HSCs produce the extracellular matrix which forms the scar tissue seen in fibrosis.[16,388] Collagen is one of the key components of this extracellular matrix, and can increase risk of progressing to cirrhosis. Endothelin increases both HSC proliferation and activation.[429] This thought to be modulated by ETA receptors. Indeed, Cho et al. found that an ETA receptor antagonist blocked the formation and deposition of collagen in rats with liver fibrosis.[110] This may explain a portion of the increased fibrosis risk in carriers of *EDNRA* risk variants associated with higher ETA receptor activity.[388] This forms a vicious cycle, as when hepatic stellate cells are activated; such as in NASH or fibrosis; secretion of ET1 is increased.[391]

The combined effect of these mechanisms and processes likely explain the increased NASH and fibrosis in those carrying genetic variants associated with higher ETR activity. This

suggests the use of drugs which interrupt and reduce the effects of ET1 this may have a role in fighting NASH, fibrosis and cirrhosis on several fronts. Endothelin receptor antagonists (ETRAs) are a class designed to down regulate endothelin receptors. These drugs have been used to treat pulmonary arterial hypertension primarily.[390] They have been investigated as potential treatments for a number of other conditions including renal failure, sickle cell disease, and cancer.[430–432]

Studies have investigated the use of ETRAs for the treatment of NASH and fibrosis.[107,391] Studies on rats have yielded promising results for the treatment of cirrhosis. Cavasin et al. found that antagonism of ETA receptors reduced hepatic portal blood pressure, and suggested a selective ETA receptor antagonist should be more effective than a dual antagonist, so as to preserve the beneficial effects of ETB receptor activation. (relaxation of sinusoids and systemic vasodilation) [433] These results were mirrored by De Gottardi et a., as they found only ETA receptors to reduce portal hypertension.[434]

Despite past research suggesting a therapeutic role of ETRAs for NASH, and the success some have had in therapy for a number of conditions, serious negative side effects have been observed. Cases of liver toxicity have been found in patients being treated with ETRAs, some of which have been fatal.[435] This is another aspect of ETRA which requires further research to fully understand the effects of the drugs and provide safe and effective treatment.

The results of this study may have clinical relevance through stratification of patients, should a therapy based on ETRAs be approved. The patients who carry the *EDNRA* risk variants and therefore have greater endothelin receptor activity, may perhaps benefit more from ETRAs than those with the wild type allele. This is an opportunity for applied personalised medicine.

6.5.1.1 Limitations and Future Work

The major limitation of this study is lack of experimental power, due to low numbers of individuals diagnosed with NASH and related phenotypes in medical records. The under diagnosis of NAFLD and related phenotypes in clinical settings is discussed previously in this thesis. The phenotypes NASH, fibrosis and NAFLD hospitalisation each had fewer than 100 cases in patients who had been genotyped in both GoDARTS and SHARE. This makes the detection of genetic effects extremely underpowered, especially with binary outcomes. As discussed in previous chapters, many patients in the GoDARTS and SHARE cohorts are likely to have had undiagnosed NASH, fibrosis and cirrhosis, which reduced effect size and the statistical significance of findings. The same issue was found for other, continuous phenotypes including FIB4, FLI, APRI and GGT.

A number of results in GoDARTS were not statistically significant when replicated in SHARE, though had the same direction of effect and similar effect sizes. The results of this current study are not strong enough to conclusively demonstrate the role of *EDN1* and *EDNRA* variants in NASH and fibrosis, but the effects seen are concurrent with previous research and are biologically plausible.

Multiple testing may have been an issue in this study. Each variant was tested in models with a number of NAFLD related phenotypes. This may have increased the number of significant results simply by chance. Techniques such as Bonferonni can be used to avoid this, but can blunt the sensitivity of analysis.

To take this research further, replication and further analyses in a larger cohort with access to high quality phenotype data is necessary. The challenge of diagnosing NAFLD causes a large underdiagnosis of NAFLD in clinical settings, which makes this kind of data rare.[25] Further to this, retrospective cohort studies investigating liver damage outcomes in those who were

prescribed ETRAs for pulmonary hypertension may be an effective method of using existing data to assess the efficacy of ETRAs for liver damage.

6.6 Conclusion

To conclude, we have shown that variants in *EDNRA* and *EDN1* which are known to associate with upregulated endothelin receptor activity have significant effects on NASH, NAFLD hospitalisation, and indexes of fibrosis. Given the association of these variants with the development of atherosclerosis and heart failure with preserved ejection fraction, these data in NASH raise the possibility that endothelin receptor antagonism may target multiple underlying pathologies. Genetic profiling may also allow for targeted therapy in patients with established NASH and fibrosis.

7 Discussion

7.1 Summary of Findings

7.1.1 NAFLD Phenotype

A NAFLD phenotype based on two raised ALT measurements at least 3 months apart, in the absence of other causes of liver disease was developed. A number of ALT thresholds were discussed, and those suggested by Prati et al. were (>19U/L for women, >30U/L for men) selected as there is good evidence that these are the upper limits of ALT levels in healthy individuals.[180] Measures were at least 3 months apart to remove the possibility of acute liver damage (e.g. drug induced) being classified as NAFLD. This definition was validated against NAFLD diagnoses from EHRs, and was 96% sensitive in GoDARTS. It correlated well with other known NAFLD risk factors such as BMI, age, T2DM and cholesterol. The established NAFLD risk SNP rs738409 was strongly associated with this NAFLD phenotype, providing more validation.

Individuals with NAFLD were found to have increased morbidity compared with those without NAFLD, as they had more hospital admissions, even when liver related admissions were excluded. Those with NAFLD also lived shorter lives, also when liver related death was excluded. This demonstrated the established underdiagnosis of NAFLD in clinical settings, as well as the effect of NAFLD on extrahepatic morbidity and mortality.[25] This has been explored in previous literature, with cardiovascular disease and cancer found to contribute significantly to these associations.[58,59] The link between NAFLD and cancer was investigated in detail in the following chapter.

The development of an accurate and practical method for the diagnosis of NAFLD in large retrospective cohorts is an important step in the field of NAFLD research. This method of NAFLD diagnosis could allow analysis of NAFLD epidemiology, pathology and genetic

determinants in cohorts which previously lacked a suitable NAFLD definition. As NAFLD diagnosis is so challenging, as simple and commonly measured definition such as this may allow analysis with many more cases compared with biopsy based definitions for example.

7.1.2 NAFLD and Cancer

The aim of this chapter was to analyse the relationship between cancer incidence and death, and NAFLD. We found that NAFLD was significantly associated with increased risk of cancer incidence in GoDARTS, SHARE and Tayside and Fife Diabetic cohort (T&F). BMI was associated with increased cancer incidence in an unadjusted model. When NAFLD was accounted for, BMI was no longer associated with cancer incidence. The same results was found when analysis was limited to previously reported BMI related cancers.[221] This result was consistent with another study into NAFLD and cancer, where Allen et al. found BMI made little to no difference to cancer risk in those without NAFLD.[59] These results combined suggest previous results linking BMI to cancer have been driven by NAFLD. NAFLD was significantly associated with increases in specific cancer incidence, including prostate, breast, colon, lung and liver cancers. In GoDARTS, the common NAFLD risk variant *PNPLA3* was significantly associated with increased cancer incidence, further validating the results of this study.

We also found NAFLD increased risk of cancer death in GoDARTS, SHARE and T&F. Cancer was found to be the predominant cause of early death in those with NAFLD, as when cancer deaths were excluded, NAFLD was not significantly associated with age of death. In GoDARTS, individuals with NAFLD who did not have T2DM, and did not have cancer at any point in their lives, lived lives of similar length to those who did not have NAFLD.

Similar findings were made by Simon et al. in a large Swedish cohort where extrahepatic cancer was the predominant cause of early death in NAFLD.[60]

These findings suggest that cancer is a major part of the epidemiology of NAFLD. This may also give clues as to the pathogenesis of certain cancers, and help the prevention of cancers through interventions against NAFLD. NAFLD is associated with numerous pathways which may influence cancer development. Hyperinsulinemia has been linked to a number of cancers, and is highly prevalent in those with NAFLD.[236,436] The pro-inflammatory state created by NAFLD may cause increased cancer risk, as increases in cytokines, adipokines and lymphokines increase cell proliferation, migration and hinder apoptosis.[238,239]

7.1.3 NAFLD GWAS

GWAS analyses of NAFLD in GoDARTS and DMDSC revealed *PNPLA3* rs738409 has a significant effect on NAFLD risk in both cohorts, with similar odds ratios. Though the data and NAFLD phenotypes were slightly different, this shows that this locus is an important factor in the development of NAFLD in both populations. Variants in *ERLIN1,* a SREBP signalling regulator, were also associated with reduced NAFLD risk in GoDARTS. This association was not seen in DMDSC, likely due to insufficient statistical power.

Fatty Liver Index was analysed in GoDARTS, and a number of genetic variants were close to genome wide significance. Variants in *GGT1* were close to significance, and as GGT is included in the calculation of FLI.[148] A second GWAS of FLI was run in GoDARTS adjusted for this *GGT1* locus, so as to remove the variance in FLI caused by this. This produced stronger results, with a number of genome wide significant hits. Significant signals were seen in chromosome 3 in the *FAM19A4* and *EOGT* genes, which have roles in inflammation and metabolism respectively.[305,307] Variants in *DNAH11* were genome wide significantly associated with decreased FLI, though this locus has not previously been associated with

NAFLD or associated pathways.[304] Two variants in *TCF7L2* were genome wide significantly associated with FLI. One of these was the variant rs7903146, which is a large T2DM risk factor, and has independent effects on NAFLD risk.[254,317] No genome wide significant associations were found for Fatty Liver Grade in the DMDCS cohort.

7.1.4 NAFLD and GLP1R, GCG and GCGR Genes

GLP1R/GCGR co-agonist medications have been used to treat diabetes and obesity, and have recently been investigated as a potential treatment for NAFLD.[97,101,103] Variants in the genes associated with these receptors have been linked to their activity, and a number of relevant metabolic pathways.[349,355] We sought to investigate whether variation in these genes was associated with risk of NAFLD in the GoDARTS, SHARE and DMDSC cohorts. We found two missense variants in *GLP1R*; rs1042044 and rs6923761; were significantly associated with NAFLD in the meta analysis of GoDARTS and SHARE. Carrying a risk genotype for either one of these *GLP1R* SNPs was associated with increased risk of NAFLD. The rs140065949 variant in *GCGR* was also significantly associated with NAFLD. Statistically significant interactions were found as well, as the effect of *GCGR* rs140065949 was significant in carriers but not non-carriers of *GLP1R* risk genotypes.

These findings suggest a role of *GLP1R* and *GCGR* genes in the development of NAFLD, and complement previous research suggesting that these receptors influence NAFLD risk. These findings also agree that GLP1R/GCGR co-agonism ought to have beneficial effects on NAFLD, and that patients with different genotypes may have different drug response. They may have utility in implementing personalised medicine in those treated with GLP1R/GCGR agonists, should they be approved for treatment of NAFLD.

7.1.5 NAFLD and Endothelin Genes

Endothelin receptor agonists have been investigated as a treatment for hepatic fibrosis.[391] Endothelin increases activation and proliferation of the HSCs which produce ECM which is present in fibrosos.[428] Studies have found increased endothelin in cirrhotic livers, and increased presence of endothelin receptors on HSCs.[437] Antagonisation of the endothelin receptor by ETRA drugs has been found to reduce hepatic fibrosis.[391]

The current study aimed to investigate the role of genetic variants related to endothelin function and NAFLD in the GoDARTS, SHARE and DMDSC cohorts. We found variants in *EDN1* and *EDNRA* to be significantly associated with NASH, NAFLD hospitalisation, fibrosis and portal hypertension. The variants in *EDNRA* have been associated with increased endothelin receptor activity in previous literature.[112,113] The variant in *EDN1* was has also been linked to increased endothelin.[112] These findings are consistent with previous literature suggesting down regulation of the endothelin receptor will reduce NAFLD and fibrosis. The findings of the current study could also be useful in stratifying patients by genotype for treatment with ETRAs, as different genotypes may respond differently.

7.2 Clinical Implications of the Current Study

The findings of the current study have major implications for a large number of individuals. NAFLD is a common condition, affecting around 25.2% of adults globally.[54] The current study showed that the vast majority of patients with NAFLD have not been admitted to hospital for their condition, and are likely undiagnosed, which is consistent with previous findings.[25] Roughly 2% of patients with NAFLD has this listed as a cause of death in the GoDARTS cohort. Despite this, NAFLD patients have much more morbidity, and died on average 1.93 years younger than those without NAFLD, after adjusting for age, sex, BMI and diabetes. This is also consistent with previous reports about mortality and NAFLD.[58]

Further to this, we showed NAFLD was associated with increased cancer incidence and cancer death, and that BMI was no longer associated with cancer incidence when NAFLD was accounted for. Screening for NAFLD may allow better prediction and targeted screening for cancers. Our results demonstrated this association with a NAFLD definition based on ALT, a cheap and commonly measured biomarker. This could be applied retrospectively, with patients with raised ALTs in the past referred for screening. It may also be applied prospectively, with increased ALT testing and routine ALT measurements used to flag those who are at risk.

Knowledge of the association between NAFLD and cancer should mean that even moderate NAFLD is taken seriously and interventions to try and reduce its severity attempted more often. Though there is no recommended pharmacological treatment, lifestyle interventions such as weight loss, exercise, smoking cessation and lowering alcohol intake can improve patients' condition.[96,200,438,439]

In 2016, It was estimated to cost the USA $103 billion dollars per year, with prevalence rates still rising.[440] Given the under diagnosis and underreporting of NAFLD in clinical settings, this number is likely much higher in reality.[25] If extrahepatic outcomes associated with NAFLD are included, this figure would again rise dramatically. The direct cost of cancer healthcare in the USA is estimated to be $173 billion for example.[441]

Further to the clinical implications, these findings suggest that more intensive research into NAFLD Is required, particularly with the aim of producing an effective drug for the therapy of NAFLD and diagnostic methods. Improved diagnostic methods would improve patient care through targeted screening for cancer, but also improve NAFLD research, as a high quality NAFLD phenotype would allow effective clinical trials to be run.

The associations with NAFLD and *GLP1R/GCGR* genes, and endothelin genes give further clues as to the pathological nature of NAFLD, and may relevant to applying precision medicine in NAFLD if drugs involved in their respective pathways are approved for NAFLD. Genotyping patients has become more and more affordable, and many consider personalised medicine for all just years away, with every patient being genotyped.[442,443]

7.3 Strengths and Limitations

This thesis used data derived from EHRs and existing high quality research studies for clinical variables. Large amounts of longitudinal data were available allowing long follow-up periods with many events, which enabled adequately powered analyses to be run in most cases. Access to genetic data for a large number of these patients allowed for a number of interesting and productive genetic studies to be undertaken.

The limitations associated with a NAFLD phenotype based on elevated ALT levels have been discussed in previous sections of the current thesis. We demonstrated the accuracy of the phenotype, validating it against NAFLD diagnosed in medical records and positive controls such as CKD, and the *PNPLA3* rs738409 variant.

Observational studies have a number of limitations compared with randomised control trials.[444,445] A common issue is confounding factors. This was countered in the current study using several techniques. Analyses were adjusted for sex, age, T2DM status, and BMI where appropriate, and cancer analyses were also adjusted for smoking status. This helped mitigate the possibility of interpreting the effects these NAFLD correlates, as effects of NAFLD itself. Exclusions for a number of different liver disease causes were made when defining the NAFLD phenotype, further removing confounding factors.

Selection bias is another issue which can affect observational studies. In the case of the current study this may have been introduced when defining NAFLD cases and controls. For

example, patients with more morbidities may interact more with healthcare services, and therefore have more ALT measurements taken, giving them more of chance of having raised measurements. The longitudinal nature of the data associated with each cohort meant that each patient on average had 20 ALT measurements, and almost all had more than two measurements. The make-up of the GoDARTS cohort; primarily a T2DM research cohort; has roughly a 60% to 40% split of diabetic to non-diabetic patients, and was designed to limit selection bias of just having T2DM patients.[187] The Tayside and Fife cohort allowed replication of results in a cohort which used population level data also.

7.4 Future Work

The findings of the current thesis suggest a number of future paths for impactful research, using currently available datasets and potentially expanding to use more.

Research into cancer genetics has uncovered a number of risk genotypes, some which have been used for applied personalised medicine, such as BRCA.[446] Analysis of the genetics which predispose to cancer specifically in people with NAFLD may help to reveal the nature of the relationship between NAFLD and cancer, as well as be another opportunity for applied personalised medicine.

We found NAFLD was associated with a number of extrahepatic cancers, and previous studies have reported links to other extrahepatic conditions such as CKD.[189] NAFLD could potentially be used to predict future incidence of these conditions. Further to this, genetic predictors of NAFLD may help in the prediction of other disease, as was shown by our Mendelian randomisation analysis for NAFLD and cancer.

The GWAS analysis of FLI in GoDARTS revealed variants in the *GGT1* gene close to genome wide statistical significance. This was controlled for in a subsequent GWAS, which produced clearer and more significant results. An investigation into the effects of the *GGT1*

variant on FLI and whether a patient's genotype could be included in the FLI calculation could improve its utility as a NAFLD biomarker.

Conceptually, NAFLD as a diagnosis has its roots as an alternative explanation for liver damage which was previously thought to only be caused by alcohol consumption.[32] This nomenclature is dated and updating it to "MAFLD" has been suggested and adopted by many in the field.[34,447] It is increasingly understood that the alcohol and non-alcohol related causes of NAFLD exist together and contribute to the disease as its seen globally.[33,448] Most adults in Europe drink some amount of alcohol, with many drinking harmful levels. [449] The current study excluded those who consumed excessive alcohol, but an effective and sensitive approach to understanding the condition may be controlling for alcohol consumption instead. Binge drinking is common in Scotland, and has been found to increase risk of liver disease, even in those who do not exceed weekly limits.[448] Though research into alcohol use is often challenging, understanding the role of drinking and particularly binge drinking in NAFLD may be key to understanding the disease. Contributions from a wide variety of factors are known to affect NAFLD, and lifestyle factors in particular have not been described fully in the literature. A key feature of the SHARE cohort is the ability to contact individuals in recruit by phenotype studies, and this could be utilised to survey the lifestyle of NAFLD cases and controls.[188] This would also help the understanding of MAFLD specifically, taking into accounts all aspects of risk.

Genetic research published in high impact journals can often be Western centric in their cohorts, and thus analysis of genetic modifiers of disease susceptibility.[450–452] Many reported genetic associations differ between ethnicities and significant genetic variation exists between South Indians and Europeans. There are many genetic variants which the two populations do not share in common. The GWAS analyses conducted in the current study used genotyping platforms which were developed mainly in Western countries.[453] Access to genotypic data for

variants specific to the South Indian population provides an opportunity to investigate whether there are any ethnicity specific genetic risk factors for NAFLD. This may further aid the understanding of NAFLD in South Indian populations, and also generally in all ethnicities. It also may provide the basis for personalised medicine for South Indian patients.

7.5 Conclusion

To conclude, the current thesis shows the development and validation of a NAFLD definition based on elevated ALT levels in the absence of alternative causes of liver disease. We show increased risk of cancer incidence and death in those with NAFLD, and that cancer is the main factor in the shortened lifespans seen in patients with NAFLD. We found *PNPLA3* rs738409 was a key component of genetic risk for NAFLD in both Scottish and Indian populations, and showed significant effects of several other genetic variants. Variants in *GCGR* and *GLP1R* had significant effects on NAFLD risk, further demonstrating a role of the receptors which they code for in NAFLD, as well as providing an opportunity for personalised medicine, should *GLP1R/GCGR* co-agonists be approved for use in treatment of NAFLD. Variants previously shown to affect the activity of endothelin had detectable effects on the susceptibility to NAFLD in our cohorts, providing another possibility for the application of personalised medicine. These findings suggest patients' genotypes may influence the efficacy of endothelin receptor antagonists in the treatment of NAFLD.

8 Bibliography

1. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology 2016;64(1):73–84.

2. Mitra S, De A, Chowdhury A. Epidemiology of non-alcoholic and alcoholic fatty liver diseases. Transl Gastroenterol Hepatol.2020;1–17.

3. Scaglioni F, Marino M, Ciccia S, et al. Short-term multidisciplinary non-pharmacological intervention is effective in reducing liver fat content assessed non-invasively in patients with nonalcoholic fatty liver disease (NAFLD). Clin Res Hepatol Gastroenterol 2013;37(4):353–8.

4. Estes C, Razavi H, Loomba R, Younossi Z, Sanyal AJ. Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease. Hepatology 2018;67(1):123–33.

5. Bertot LC, Adams LA. The natural course of non-alcoholic fatty liver disease. Int J Mol Sci 2016;17(5).

6. Baffy G. MicroRNAs in nonalcoholic fatty liver disease. J Clin Med 2015;4(12):1977–88.

7. Loomba R, Sanyal AJ. The global NAFLD epidemic. Nat Rev Gastroenterol Hepatol [Internet] 2013;10(11):686–90. Available from: https://doi.org/10.1038/nrgastro.2013.171

8. Nassir F, Rector RS, Hammoud GM, Ibdah JA. Pathogenesis and Prevention of Hepatic Steatosis. Gastroenterol Hepatol (N Y) [Internet] 2015;11(3):167–75. Available from: https://pubmed.ncbi.nlm.nih.gov/27099587

9. Marra, Fabio, and Sophie Lotersztajn. "Pathophysiology of NASH: perspectives for a targeted treatment." Current pharmaceutical design 19.29 (2013): 5250-5269.

10. Diehl AM, Day C. Cause, pathogenesis, and treatment of nonalcoholic steatohepatitis. N Engl J Med 2017;377:2063–72.

11. Masarone M, Rosato V, Dallio M, et al. Role of Oxidative Stress in Pathophysiology of Nonalcoholic Fatty Liver Disease. Oxid Med Cell Longev [Internet] 2018;2018:9547613. Available from: https://pubmed.ncbi.nlm.nih.gov/29991976

12. Viganò L, Lleo A, Aghemo A. Non-alcoholic fatty liver disease, non-alcoholic steatohepatitis, metabolic syndrome and hepatocellular carcinoma-a composite scenario. Hepatobiliary Surg Nutr [Internet] 2018;7(2):130–3. Available from: https://pubmed.ncbi.nlm.nih.gov/29744343

13. Brown GT, Kleiner DE. Histopathology of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. Metabolism [Internet] 2016;65(8):1080–6. Available from: https://pubmed.ncbi.nlm.nih.gov/26775559

14. Ratziu V, Charlotte F, Heurtier A, et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. Gastroenterology 2005;128(7):1898–906.

15. Böttcher K, Pinzani M. Pathophysiology of liver fibrosis and the methodological barriers to the development of anti-fibrogenic agents. Adv Drug Deliv Rev [Internet] 2017;121:3–8. Available from: http://www.sciencedirect.com/science/article/pii/S0169409X17300728

16. Puche JE, Saiman Y, Friedman SL. Hepatic stellate cells and liver fibrosis. Compr Physiol 2011;3(4):1473–92.

17. CASTRO PCS de, ALBERTON HCP, PEDROSO MLA, MORSOLETTO DBG,

PISSAIA JUNIOR A, IVANTES CAP. Evaluation of progression of hepatic fibrosis in a group of patients with non-alcoholic fatty liver disease accompanied for 10 years. Arq Gastroenterol 2019;56(3):256–60.

18. Povero D, Busletta C, Novo E, et al. Liver fibrosis: a dynamic and potentially reversible process. HISTOLOGY AND HISTOPATHOLOGY 2010;

19. MATTEONI C, YOUNOSSI Z, GRAMLICH T, BOPARAI N, LIU Y, MCCULLOUGH A. Nonalcoholic fatty liver disease: A spectrum of clinical and pathological severity☆, ☆☆. Gastroenterology [Internet] 1999;116(6):1413–9. Available from: http://www.sciencedirect.com/science/article/pii/S0016508599705068

20. Suk KT, Kim DJ. Staging of liver fibrosis or cirrhosis: The role of hepatic venous pressure gradient measurement. World J Hepatol 2015;7(3):607–15.

21. Hanna RF, Aguirre DA, Kased N, Emery SC, Peterson MR, Sirlin CB. Cirrhosis-associated hepatocellular nodules: correlation of histopathologic and MR imaging features. Radiographics 2008;28(3):747–69.

22. Margini C, Dufour JF. The story of HCC in NAFLD : from epidemiology , across pathogenesis , to prevention and treatment. Liver International. 2016;(August 2015):317–24.

23. Desai A, Sandhu S, Lai J-P, Sandhu DS. Hepatocellular carcinoma in non-cirrhotic liver: A comprehensive review. World J Hepatol [Internet] 2019;11(1):1–18. Available from: https://pubmed.ncbi.nlm.nih.gov/30705715

24. Ramakrishna G, Rastogi A, Trehanpati N, Sen B, Khosla R, Sarin SK. From cirrhosis to hepatocellular carcinoma: new molecular insights on inflammation and cellular senescence. Liver cancer [Internet] 2013;2(3–4):367–83. Available from:

https://pubmed.ncbi.nlm.nih.gov/24400224

25. Alexander M, Loomis AK, Fairburn-Beech J, et al. Real-world data reveal a diagnostic gap in non-alcoholic fatty liver disease. BMC Med 2018;16(1):1–11.

26. Ahmed Z, Ahmed U, Walayat S, et al. Liver function tests in identifying patients with liver disease. Clin Exp Gastroenterol 2018;11:301–7.

27. Clark JM, Diehl AM. Nonalcoholic Fatty Liver Disease An Underrecognized Cause of Cryptogenic Cirrhosis. JAMA. 2017;289(22):1–5.

28. Patch D, Luong TV. Biopsy of the Liver [Internet]. Sherlock's Dis. Liver Biliary Syst. 2018;39–52. Available from: https://doi.org/10.1002/9781119237662.ch3

29. Rinella ME, Loomba R, Caldwell SH, et al. Controversies in the diagnosis and management of NAFLD and NASH. Gastroenterol Hepatol 2014;10(4):219–27.

30. Mishra P, Younossi ZM. Abdominal ultrasound for diagnosis of nonalcoholic fatty liver disease (NAFLD). Am. J. Gastroenterol. 2007;102(12):2716–7.

31. Miyake T, Kumagi T, Hirooka M, et al. Metabolic markers and ALT cutoff level for diagnosing nonalcoholic fatty liver disease: a community-based cross-sectional study. J Gastroenterol 2012;47(6):696–703.

32. Ludwig J, Viggiano TR, McGill DB, Oh BJ. Nonalcoholic steatohepatitis: Mayo Clinic experiences with a hitherto unnamed disease. Mayo Clin Proc 1980;55(7):434–8.

33. Fan J-G. Epidemiology of alcoholic and nonalcoholic fatty liver disease in China. J Gastroenterol Hepatol 2013;28 Suppl 1:11–7.

34. Eslam M, Sanyal AJ, George J, et al. MAFLD: A Consensus-Driven Proposed

Nomenclature for Metabolic Associated Fatty Liver Disease. Gastroenterology [Internet] 2020;Available from: http://www.sciencedirect.com/science/article/pii/S0016508520301712

35. Buzzetti E, Pinzani M, Tsochatzis EA. The multiple-hit pathogenesis of non-alcoholic fatty liver disease (NAFLD). Metabolism [Internet] 2016;65(8):1038–48. Available from: http://dx.doi.org/10.1016/j.metabol.2015.12.012

36. Danforth Jr E. Diet and obesity. Am J Clin Nutr 1985;

37. Mirza MS. Obesity, Visceral Fat, and NAFLD: Querying the Role of Adipokines in the Progression of Nonalcoholic Fatty Liver Disease. ISRN Gastroenterol 2011;2011:1–11.

38. Bischoff SC, Barbara G, Buurman W, et al. Intestinal permeability--a new target for disease prevention and therapy. BMC Gastroenterol [Internet] 2014;14:189. Available from: https://pubmed.ncbi.nlm.nih.gov/25407511

39. Henkel A, Green RM. The unfolded protein response in fatty liver disease. Semin Liver Dis 2013;33(4):321–9.

40. Sumida Y, Niki E, Naito Y, Yoshikawa T. Involvement of free radicals and oxidative stress in NAFLD/NASH. Free Radic Res 2013;47(11):869–80.

41. Younossi ZM, Stepanova M, Negro F, et al. Nonalcoholic Fatty Liver Disease in Lean Individuals in the United States. Medicine (Baltimore) [Internet] 2012;91(6):319–27. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00005792-201211000-00003

42. Sattar N, Forrest E, Preiss D. Non-alcoholic fatty liver disease. BMJ [Internet]

2014;349. Available from: https://www.bmj.com/content/349/bmj.g4596

43. Romero-Gómez M, Zelber-Sagi S, Trenell M. Treatment of NAFLD with diet, physical activity and exercise. J Hepatol 2017;67(4):829–46.

44. Glen J, Floros L, Day C, Pryke R. Non-alcoholic fatty liver disease (NAFLD): summary of NICE guidance. BMJ [Internet] 2016;354:i4428. Available from: http://www.bmj.com/content/354/bmj.i4428.abstract

45. Pandyarajan V, Gish RG, Alkhouri N, Noureddin M. Screening for Nonalcoholic Fatty Liver Disease in the Primary Care Clinic. Gastroenterol Hepatol (N Y) [Internet] 2019;15(7):357–65. Available from: https://pubmed.ncbi.nlm.nih.gov/31391806

46. Nascimbeni F, Pais R, Bellentani S, et al. From NAFLD in clinical practice to answers from guidelines. J Hepatol [Internet] 2013;59(4):859–71. Available from: http://www.sciencedirect.com/science/article/pii/S0168827813003772

47. Caussy C. Should We Screen High-Risk Populations for NAFLD? Curr Hepatol Reports 2019;18(4):433–43.

48. Zhang E, Wartelle-Bladou C, Lepanto L, Lachaine J, Cloutier G, Tang A. Cost-utility analysis of nonalcoholic steatohepatitis screening. Eur Radiol [Internet] 2015;25(11):3282–94. Available from: https://doi.org/10.1007/s00330-015-3731-2

49. Younossi ZM. Non-alcoholic fatty liver disease – A global public health perspective. J Hepatol [Internet] 2019;70(3):531–44. Available from: https://doi.org/10.1016/j.jhep.2018.10.033

50. Rowe IA. Too much medicine: overdiagnosis and overtreatment of non-alcoholic fatty liver disease. lancet Gastroenterol Hepatol 2018;3(1):66–72.

51. Malnick S. 5 Non-alcoholic fatty liver disease (NAFLD) -underdiagnosed but

overtreated. BMJ Evidence-Based Med [Internet] 2019;24(Suppl 2):A54 LP-A55. Available from: http://ebm.bmj.com/content/24/Suppl_2/A54.2.abstract

52. Romeo S, Kozlitina J, Xing C, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nat Genet 2008;40(12):1461–5.

53. Losekann A, Weston AC, De Mattos AA, et al. Non-alcoholic steatohepatitis (NASH): Risk factors in morbidly obese patients. Int J Mol Sci 2015;16(10):25552–9.

54. Bellentani S, Scaglioni F, Marino M, Bedogni G. Epidemiology of non-alcoholic fatty liver disease. Clin Liver Dis [Internet] 2016;7(5):106–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20460905

55. Younossi ZM, Golabi P, de Avila L, et al. The global epidemiology of NAFLD and NASH in patients with type 2 diabetes: A systematic review and meta-analysis. J Hepatol [Internet] 2019;71(4):793–801. Available from: http://www.sciencedirect.com/science/article/pii/S0168827819303939

56. Day CP. Natural history of NAFLD: remarkably benign in the absence of cirrhosis. Gastroenterology 2005;129(1):375–8.

57. Ekstedt M, Hagström H, Nasr P, et al. Fibrosis stage is the strongest predictor for disease-specific mortality in NAFLD after up to 33 years of follow-up. Hepatology 2015;61(5):1547–54.

58. Allen AM, Therneau TM, Larson JJ, Coward A, Somers VK, Kamath PS. Nonalcoholic fatty liver disease incidence and impact on metabolic burden and death: A 20 year-community study. Hepatology 2018;67(5):1726–36.

59. Allen AM, Hicks SB, Mara KC, Larson JJ, Therneau TM. The Risk of Incident Extrahepatic Cancers is higher in Nonalcoholic Fatty Liver Disease than Obesity - a

Longitudinal Cohort Study. J Hepatol 2019;

60. Simon TG, Roelstraete B, Khalili H, Hagström H, Ludvigsson JF. Mortality in biopsy-confirmed nonalcoholic fatty liver disease: results from a nationwide cohort. Gut [Internet] 2020;gutjnl-2020-322786. Available from: http://gut.bmj.com/content/early/2020/10/18/gutjnl-2020-322786.abstract

61. Lattuada G, Ragogna F, Perseghin G. Why does NAFLD predict type 2 diabetes? Curr Diab Rep 2011;11(3):167–72.

62. Bertolotti M, Lonardo A, Mussi C, et al. Nonalcoholic fatty liver disease and aging: Epidemiology to management. World J Gastroenterol 2014;20(39):14185–204.

63. Than NN, Newsome PN. A concise review of non-alcoholic fatty liver disease. Atherosclerosis [Internet] 2015;239(1):192–202. Available from: http://dx.doi.org/10.1016/j.atherosclerosis.2015.01.001

64. Manuscript A, Implications C. Obesity and Nonalcoholic Fatty Liver Disease: Biochemical, Metabolic and Clinical Implications NIH Public Access. Hepatology 2010;51(2):679–89.

65. Fabbrini E, Sullivan S, Klein S. Obesity and nonalcoholic fatty liver disease: Biochemical, metabolic, and clinical implications. Hepatology [Internet] 2010;51(2):679–89. Available from: https://doi.org/10.1002/hep.23280

66. Pang Q, Zhang J-Y, Song S-D, et al. Central obesity and nonalcoholic fatty liver disease risk after adjusting for body mass index. World J Gastroenterol [Internet] 2015;21(5):1650–62. Available from: https://pubmed.ncbi.nlm.nih.gov/25663786

67. Ghaemi A, Taleban FA, Hekmatdoost A, et al. How Much Weight Loss is Effective on Nonalcoholic Fatty Liver Disease? Hepat Mon [Internet] 2013;13(12):e15227–e15227.

Available from: https://pubmed.ncbi.nlm.nih.gov/24358045

68. Keating SE, Hackett DA, George J, Johnson NA. Exercise and non-alcoholic fatty liver disease : A systematic review and meta-analysis. J Hepatol [Internet] 2012;57(1):157–66. Available from: http://dx.doi.org/10.1016/j.jhep.2012.02.023

69. Gerber L, Otgonsuren M, Mishra A, et al. Non-alcoholic fatty liver disease (NAFLD) is associated with low level of physical activity: a population-based study. Aliment Pharmacol Ther 2012;36(8):772–81.

70. VanWagner LB, Armstrong MJ. Lean NAFLD: A not so benign condition? Hepatol Commun 2018;2(1):5–8.

71. Koehler EM, Schouten JNL, Hansen BE, et al. Prevalence and risk factors of non-alcoholic fatty liver disease in the elderly: results from the Rotterdam study. J Hepatol 2012;57(6):1305–11.

72. Lonardo A, Lombardini S, Scaglioni F, et al. Fatty liver, carotid disease and gallstones: a study of age-related associations. World J Gastroenterol WJG 2006;12(36):5826.

73. Tilg H, Moschen AR, Roden M. NAFLD and diabetes mellitus. Nat Rev Gastroenterol Hepatol [Internet] 2017;14(1):32–42. Available from: https://doi.org/10.1038/nrgastro.2016.147

74. Younossi ZM, Gramlich T, Matteoni CA, Boparai N, McCullough AJ. Nonalcoholic fatty liver disease in patients with type 2 diabetes. Clin Gastroenterol Hepatol 2004;2(3):262–5.

75. Loomba R, Abraham M, Unalp A, et al. Association between diabetes, family history of diabetes, and risk of nonalcoholic steatohepatitis and fibrosis. Hepatology [Internet] 2012;56(3):943–51. Available from: https://doi.org/10.1002/hep.25772

76. Smith BW, Adams LA. Nonalcoholic fatty liver disease and diabetes mellitus: pathogenesis and treatment. Nat Rev Endocrinol [Internet] 2011;7(8):456–65. Available from: https://doi.org/10.1038/nrendo.2011.72

77. Anstee QM, Targher G, Day CP. Progression of NAFLD to diabetes mellitus, cardiovascular disease or cirrhosis. Nat Rev Gastroenterol Hepatol [Internet] 2013;10(6):330–44. Available from: https://doi.org/10.1038/nrgastro.2013.41

78. Adams L. NAFLD as a risk factor for the development of diabetes and the metabolic syndrome: an eleven-year follow-up study. American Journal of Gastroenterology, 104(4), 861-867.

79. Howard B V, Ruotolo G, Robbins DC. Obesity and dyslipidemia. Endocrinol Metab Clin North Am 2003;32(4):855.

80. Souza MR de A, Diniz M de FF de M, Medeiros-Filho JEM de, Araújo MST de. Metabolic syndrome and risk factors for non-alcoholic fatty liver disease. Arq Gastroenterol 2012;49(1):89–96.

81. Bano A, Chaker L, Plompen EPC, et al. Thyroid function and the risk of nonalcoholic fatty liver disease: the Rotterdam Study. J Clin Endocrinol Metab 2016;101(8):3204–11.

82. Duseja A, Chalasani N, Nash NÁ, Cryptogenic Á. Epidemiology and risk factors of nonalcoholic fatty liver disease (NAFLD) Hepatology International. 2014;7(2013).

83. Sookoian S, Pirola CJ. Meta-analysis of the influence of I148M variant of patatin-like phospholipase domain containing 3 gene (PNPLA3) on the susceptibility and histological severity of nonalcoholic fatty liver disease. Hepatology [Internet] 2011;53(6):1883–94. Available from: https://doi.org/10.1002/hep.24283

84. Yuan S, Liu H, Yuan D, et al. PNPLA3 I148M mediates the regulatory effect of NF-kB on inflammation in PA-treated HepG2 cells. J Cell Mol Med [Internet] 2020;24(2):1541–52. Available from: https://doi.org/10.1111/jcmm.14839

85. Dongiovanni P, Donati B, Fares R, et al. PNPLA3 I148M polymorphism and progressive liver disease. World J Gastroenterol [Internet] 2013;19(41):6969–78. Available from: https://pubmed.ncbi.nlm.nih.gov/24222941

86. Dong XC. PNPLA3—A Potential Therapeutic Target for Personalized Treatment of Chronic Liver Disease [Internet]. Front. Med. . 2019;6:304. Available from: https://www.frontiersin.org/article/10.3389/fmed.2019.00304

87. Dongiovanni P, Romeo S, Valenti L. Genetic Factors in the Pathogenesis of Nonalcoholic Fatty Liver and Steatohepatitis. BioMed research international. 2015;2015(Vldl).

88. Anstee QM, Day CP. The genetics of NAFLD. Nat Rev Gastroenterol Hepatol 2013;10(11):645–55.

89. Gouda W, Ashour E, Shaker Y, Ezzat W. MTP genetic variants associated with non-alcoholic fatty liver in metabolic syndrome patients. Genes Dis [Internet] 2017;4(4):222–8. Available from: https://pubmed.ncbi.nlm.nih.gov/30258926

90. Eslam M, Valenti L, Romeo S, et al. Heritability of Nonalcoholic Fatty Liver Disease. Metabolism [Internet] 2018;131(2 SUPPL. 1):1585–92. Available from: http://dx.doi.org/10.1016/j.metabol.2015.12.012

91. Nobili V, Donati B, Panera N, et al. A 4-polymorphism risk score predicts steatohepatitis in children with nonalcoholic fatty liver disease. J Pediatr Gastroenterol Nutr 2014;58(5):632–6.

92. Vespasiani-Gentilucci U, Gallo P, Dell'Unto C, Volpentesta M, Antonelli-Incalzi R, Picardi A. Promoting genetics in non-alcoholic fatty liver disease: Combined risk score through polymorphisms and clinical variables. World J Gastroenterol [Internet] 2018;24(43):4835–45. Available from: https://pubmed.ncbi.nlm.nih.gov/30487694

93. Leon-Mimila P, Vega-Badillo J, Gutierrez-Vidal R, et al. A genetic risk score is associated with hepatic triglyceride content and non-alcoholic steatohepatitis in Mexicans with morbid obesity. Exp Mol Pathol 2015;98(2):178–83.

94. Eslam M, George J. Genetic Insights for Drug Development in NAFLD. Trends Pharmacol Sci [Internet] 2019;40(7):506–16. Available from: http://www.sciencedirect.com/science/article/pii/S0165614719300963

95. Abul-Husn NS, Cheng X, Li AH, et al. A Protein-Truncating *HSD17B13* Variant and Protection from Chronic Liver Disease. N Engl J Med [Internet] 2018;378(12):1096–106. Available from: http://www.nejm.org/doi/10.1056/NEJMoa1712191

96. Zelber-Sagi S, Buch A, Yeshua H, et al. Effect of resistance training on non-alcoholic fatty-liver disease a randomized-clinical trial. World J Gastroenterol 2014;20(15):4382–92.

97. Patel V, Joharapurkar A, Kshirsagar S, et al. Coagonist of glucagon-like peptide-1 and glucagon receptors ameliorates nonalcoholic fatty liver disease. Can J Physiol Pharmacol 2018;96(6):587–96.

98. Mancina RM, Dongiovanni P, Petta S, et al. The MBOAT7-TMC4 Variant rs641738 Increases Risk of Nonalcoholic Fatty Liver Disease in Individuals of European Descent. Gastroenterology 2016;150(5):1219-1230e6.

99. Chatterjee A, Basu A, Chowdhury A, Das K. Comparative analyses of genetic risk

prediction methods reveal extreme diversity of genetic predisposition to nonalcoholic fatty liver disease (NAFLD) among ethnic populations of India. 2015;94(1):105–13.

100. Kanth VVR, Sasikala M, Rao PN, Steffie Avanthi U, Rao KR, Nageshwar Reddy D. Pooled genetic analysis in ultrasound measured non-alcoholic fatty liver disease in Indian subjects: A pilot study. World J Hepatol [Internet] 2014;6(6):435–42. Available from: https://pubmed.ncbi.nlm.nih.gov/25018854

101. Boland ML, Laker RC, Mather K, et al. Resolution of NASH and hepatic fibrosis by the GLP-1R and GCGR dual-agonist cotadutide via modulating mitochondrial function and lipogenesis. Nat Metab 2020;2(5):413–31.

102. FAROOQ G, JONES BEN, MINNION JS, BLOOM SR. Effect of Biased GLP-1/Glucagon Receptor Co-agonists on Insulin Secretion. Diabetes [Internet] 2018;67(Supplement 1). Available from: https://diabetes.diabetesjournals.org/content/67/Supplement_1/1100-P

103. Day JW, Gelfanov V, Smiley D, et al. Optimization of co-agonism at GLP-1 and glucagon receptors to safely maximize weight reduction in DIO-rodents. Pept Sci [Internet] 2012;98(5):443–50. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/bip.22072

104. Lin C-H, Lee Y-S, Huang Y-Y, Hsieh S-H, Chen Z-S, Tsai C-N. Polymorphisms of GLP-1 receptor gene and response to GLP-1 analogue in patients with poorly controlled type 2 diabetes. J Diabetes Res 2015;2015.

105. Neumann UH, Ho JSS, Mojibian M, Covey SD, Charron MJ, Kieffer TJ. Glucagon receptor gene deletion in insulin knockout mice modestly reduces blood glucose and ketones but does not promote survival. Mol Metab 2016;5(8):731–6.

106. Maslak E, Gregorius A, Chlopicki S. Liver sinusoidal endothelial cells (LSECs) function and NAFLD; NO-based therapy targeted to the liver. Pharmacol Reports 2015;67(4):689–94.

107. Okamoto T, Koda M, Miyoshi K, Onoyama T, Kishina M. Antifibrotic effects of ambrisentan , an endothelin-A receptor antagonist , in a non-alcoholic steatohepatitis mouse model. World J Hepatol. 2016;8(22):933–41.

108. Motte S, Mcentee K, Naeije R. Endothelin receptor antagonists. Pharmacology & therapeutics. 2006;110:386–414.

109. Ling L, Kuc RE, Maguire JJ, et al. Comparison of endothelin receptors in normal versus cirrhotic human liver and in the liver from endothelial cell-speci fi c ET B knockout mice. Life Sci [Internet] 2012;91(13–14):716–22. Available from: http://dx.doi.org/10.1016/j.lfs.2012.02.003

110. Cho J-J, Hocher B, Herbst H, et al. An oral endothelin-A receptor antagonist blocks collagen synthesis and deposition in advanced rat liver fibrosis. Gastroenterology [Internet] 2000;118(6):1169–78. Available from: http://www.sciencedirect.com/science/article/pii/S0016508500703702

111. Wells RG. Cellular sources of extracellular matrix in hepatic fibrosis. Clin Liver Dis 2008;12(4):759–68, viii.

112. Calabrò P, Limongelli G, Maddaloni V, et al. Analysis of endothelin-1 and endothelin-1 receptor A gene polymorphisms in patients with pulmonary arterial hypertension. Intern Emerg Med [Internet] 2012;7(5):425–30. Available from: https://doi.org/10.1007/s11739-011-0643-2

113. Zhang L, Sui R, Zhang L. Effect of SNP Polymorphisms of EDN1 , EDNRA , and

EDNRB Gene on Ischemic Stroke. Cell biochemistry and biophysics 2014;233–9.

114. Gupta RM, Hadaya J, Trehan A, et al. A Genetic Variant Associated with Five Vascular Diseases Is a Distal Regulator of Endothelin-1 Gene Expression. Cell [Internet] 2017;170(3):522-533.e15. Available from: http://www.sciencedirect.com/science/article/pii/S0092867417307687

115. Dichgans M, Malik R, König IR, et al. Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. Stroke [Internet] 2014;45(1):24–36. Available from: https://pubmed.ncbi.nlm.nih.gov/24262325

116. Iqbal U, Perumpail BJ, Akhtar D, Kim D, Ahmed A. The epidemiology, risk profiling and diagnostic challenges of nonalcoholic fatty liver disease. Medicines 2019;6(1):41.

117. Karim MF, Al-Mahtab M, Rahman S, Debnath CR. Non-alcoholic Fatty Liver Disease (NAFLD)--A Review. Mymensingh Med J [Internet] 2015;24(4):873—880. Available from: http://europepmc.org/abstract/MED/26620035

118. Newton JL. Systemic Symptoms in Non-Alcoholic Fatty Liver Disease. Digestive diseases. 2010;214–9.

119. Ii WBS. Nonalcoholic Fatty Liver Disease (NAFLD): A Comprehensive Review DISORDERS OF THE LIVER DIAGNOSIS OF FATTY LIVER AND. World J Gastroenterol. 2004;27–41.

120. Adams LA, Lymp JF, St. Sauver J, et al. The natural history of nonalcoholic fatty liver disease: A population-based cohort study. Gastroenterology 2005;129(1):113–21.

121. Tsai Y-L, Liu C-W, Huang S-F, et al. Urinary fatty acid and retinol binding protein-4 predict CKD progression in severe NAFLD patients with hypertension: 4-year study

with clinical and experimental approaches. Medicine (Baltimore) 2020;99(2).

122. Bennett CM, Guo M, Dharmage SC. HbA 1c as a screening tool for detection of Type 2 diabetes : a systematic review. 2007;333–43.

123. Bennett CM, Guo M, Dharmage SC. HbA1c as a screening tool for detection of type 2 diabetes: a systematic review. Diabet Med 2007;24(4):333–43.

124. Younossi ZM, Loomba R, Anstee QM, et al. Diagnostic modalities for nonalcoholic fatty liver disease, nonalcoholic steatohepatitis, and associated fibrosis. Hepatology 2018;68(1):349–60.

125. Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology 2018;67(1):328–57.

126. Sohrabpour AA, Mohamadnejad M, Malekzadeh R. Alimentary Pharmacology and Therapeutics Review article : the reversibility of cirrhosis. Alimentary Pharmacology and Therapeutics 2012;(September).

127. Ghaemi A, Taleban FA, Hekmatdoost A, et al. How Much Weight Loss is Effective on Nonalcoholic Fatty Liver Disease ? Hepat Mon. 2013;13(12).

128. Ray K. NAFLD—the next global epidemic. Nat Publ Gr 2013;10(November):2013.

129. Byrne CD, Targher G. What's new in NAFLD pathogenesis, biomarkers and treatment? Nat Rev Gastroenterol Hepatol 2020;17(2):70–1.

130. Bambha K, Belt P, Abraham M, et al. Ethnicity and nonalcoholic fatty liver disease. Hepatology 2012;55(3):769–80.

131. Schindhelm RK, Diamant M, Dekker JM, Tushuizen ME, Teerlink T, Heine RJ.

Alanine aminotransferase as a marker of non-alcoholic fatty liver disease in relation to type 2 diabetes mellitus and cardiovascular disease. Diabetes Metab Res Rev 2006;22(6):437–43.

132. Adams LA, Feldstein AE. Non-invasive diagnosis of nonalcoholic fatty liver and nonalcoholic steatohepatitis. J Dig Dis 2011;12(1):10–6.

133. Rockey DC, Caldwell SH, Goodman ZD, Nelson RC, Smith AD. AASLD position paper: Liver biopsy. Hepatology 2008;(3):NA-NA.

134. Piccinino F, Sagnelli E, Pasquale G, Giusti G. Complications following percutaneous liver biopsy. A multicentre retrospective study on 68,276 biopsies. J Hepatol 1986;2(2):165–73.

135. Hashimoto E, Farrell GC. Will non-invasive markers replace liver biopsy for diagnosing and staging fibrosis in non-alcoholic steatohepatitis? J Gastroenterol Hepatol 2009;24(4):501–3.

136. Yoneda M, Yoneda M, Mawatari H, et al. Noninvasive assessment of liver fibrosis by measurement of stiffness in patients with nonalcoholic fatty liver disease (NAFLD). Dig Liver Dis 2008;40(5):371–8.

137. Eddowes PJ, McDonald N, Davies N, et al. Utility and cost evaluation of multiparametric magnetic resonance imaging for the assessment of non-alcoholic fatty liver disease. Aliment Pharmacol Ther 2018;47(5):631–44.

138. Khov N, Sharma A, Riley TR. Bedside ultrasound in the diagnosis of nonalcoholic fatty liver disease. World J Gastroenterol WJG 2014;20(22):6821.

139. Mahale AR, Prabhu SD. Clinical relevance of reporting fatty liver on ultrasound in asymptomatic patients during routine health checkups. 2018;

140. Hernaez R, Lazo M, Bonekamp S, et al. Diagnostic Accuracy and Reliability of Ultrasonography for the Detection of Fatty Liver: A Meta-Analysis. 2011;1082–90.

141. Paige JS, Bernstein GS, Wolfson T, et al. Gastrointestinal Imaging • Original Research. 2017;(May):168–77.

142. Hamer OW, Aguirre DA, Casola G, Lavine JE, Woenckhaus M, Sirlin CB. Fatty liver: imaging patterns and pitfalls. Radiographics 2006;26(6):1637–53.

143. Saadeh S, Younossi ZM, Remer EM, et al. The utility of radiological imaging in nonalcoholic fatty liver disease. Gastroenterology 2002;123(3):745–50.

144. Willcox S, Seddon M, Dunn S, Edwards RT, Pearse J, Tu J V. Measuring and reducing waiting times: a cross-national comparison of strategies. Health Aff (Millwood) 2007;26(4):1078–87.

145. Wong VW-S, Adams LA, de Ledinghen V, Wong GL-H, Sookoian S. Noninvasive biomarkers in NAFLD and NASH - current progress and future promise. Nat Rev Gastroenterol Hepatol 2018;15(8):461–78.

146. Hoekstra LT, de Graaf W, Nibourg GAA, et al. Physiological and biochemical basis of clinical liver function tests: a review. Ann Surg 2013;257(1):27–36.

147. Kleiner DE, Brunt EM, Van Natta M, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 2005;41(6):1313–21.

148. Bedogni G, Bellentani S, Miglioli L, et al. The fatty liver index: A simple and accurate predictor of hepatic steatosis in the general population. BMC Gastroenterol 2006;6:1–7.

149. Dehnavi Z, Razmpour F, Naseri MB, et al. Fatty liver index (FLI) in predicting non-alcoholic fatty liver disease (NAFLD). Hepat Mon 2018;18(2).

150. Lee J, Kim D, Jung H, et al. Hepatic steatosis index : A simple screening tool reflecting nonalcoholic fatty liver disease. Dig Liver Dis [Internet] 2010;42(7):504–9. Available from: http://dx.doi.org/10.1016/j.dld.2009.08.002

151. Yilmaz Y, Yonal O, Kurt R, Bayrak M, Aktas B, Ozdogan O. Noninvasive assessment of liver fibrosis with the aspartate transaminase to platelet ratio index (APRI): Usefulness in patients with chronic liver disease: APRI in chronic liver disease. Hepat Mon [Internet] 2011;11(2):103–6. Available from: https://pubmed.ncbi.nlm.nih.gov/22087126

152. Lin Z-H, Xin Y-N, Dong Q-J, et al. Performance of the aspartate aminotransferase-to-platelet ratio index for the staging of hepatitis C-related fibrosis: an updated meta-analysis. Hepatology 2011;53(3):726–36.

153. Sterling RK, Lissen E, Clumeck N, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. Hepatology [Internet] 2006;43(6):1317–25. Available from: https://doi.org/10.1002/hep.21178

154. Giannini EG, Testa R, Savarino V. Liver enzyme alteration: a guide for clinicians. C Can Med Assoc J = J l'Association medicale Can 2005;172(3):367–79.

155. Pratt DS, Kaplan MM. Evaluation of abnormal liver-enzyme results in asymptomatic patients. N Engl J Med 2000;342(17):1266–71.

156. Limdi JK, Hyde GM. Evaluation of abnormal liver function tests. Postgrad Med J 2003;79(932):307–12.

157. Tomizawa M, Kawanabe Y, Shinozaki F, et al. Elevated levels of alanine transaminase and triglycerides within normal limits are associated with fatty liver. Exp Ther Med 2014;8(3):759–62.

158. Martin-Rodriguez JL, Gonzalez-Cantero J, Gonzalez-Cantero A, Arrebola JP, Gonzalez-Calvin JL. Diagnostic accuracy of serum alanine aminotransferase as biomarker for nonalcoholic fatty liver disease and insulin resistance in healthy subjects, using 3T MR spectroscopy. Med (United States) 2017;96(17).

159. Verma S, Jensen D, Hart J, Mohanty SR. Predictive value of ALT levels for non-alcoholic steatohepatitis (NASH) and advanced fibrosis in non-alcoholic fatty liver disease (NAFLD). Liver Int 2013;33(9):1398–405.

160. Serper M, Vujkovic M, Kaplan DE, et al. Validating a non-invasive, ALT-based non-alcoholic fatty liver phenotype in the million veteran program. PLoS One 2020;15(8):e0237430.

161. Phillips ML, Boase S, Wahlroos S, et al. Associates of change in liver fat content in the morbidly obese after laparoscopic gastric banding surgery. Diabetes, Obes Metab [Internet] 2008;10(8):661–7. Available from: https://doi.org/10.1111/j.1463-1326.2007.00793.x

162. Pacifico L, Celestre M, Anania C, Paolantonio P, Chiesa C, Laghi A. MRI and ultrasound for hepatic fat quantification : relationships to clinical and metabolic characteristics of pediatric nonalcoholic fatty liver disease. 2007;542–7.

163. Maximos M, Bril F, Sanchez PP, et al. The Role of Liver Fat and Insulin Resistance as Determinants of Plasma Aminotransferase Elevation in Nonalcoholic Fatty Liver Disease. 2014;153–60.

164. Wong CA, Araneta MRG, Barrett-Connor E, Alcaraz J, Castañeda D, Macera C. Probable NAFLD, by ALT levels, and diabetes among Filipino-American Women. Diabetes Res Clin Pract 2008;79(1):133–40.

165. Yoo JS, Lee SY, Kim KN, Yoo SM, Sung EJ, Yim JE. Relationship between insulin resistance and serum alanine aminotransferase as a surrogate of NAFLD (nonalcoholic fatty liver disease) in obese Korean children. Diabetes Res Clin Pract 2008;

166. Mazzotti A, Caletti MT, Brodosi L, et al. An internet-based approach for lifestyle changes in patients with NAFLD: Two-year effects on weight loss and surrogate markers. J Hepatol 2018;69(5):1155–63.

167. DeVore S, Kohli R, Lake K, et al. A multidisciplinary clinical program is effective in stabilizing BMI and reducing transaminase levels in pediatric patients with NAFLD. J Pediatr Gastroenterol Nutr 2013;57(1):119.

168. Sorbi D, Boynton J, Lindor KD. The ratio of aspartate aminotransferase to alanine aminotransferase: potential value in differentiating nonalcoholic steatohepatitis from alcoholic liver disease. Am J Gastroenterol 1999;94(4):1018–22.

169. Maher JJ. Exploring alcohol's effects on liver function. Alcohol Health Res World 1997;21(1):5–12.

170. Cohen JA, Kaplan MM. The SGOT/SGPT ratio--an indicator of alcoholic liver disease. Dig Dis Sci 1979;24(11):835–8.

171. Heard K, Green JL, Bailey JE, Bogdan GM, Dart RC. A randomized trial to determine the change in alanine aminotransferase during 10 days of paracetamol (acetaminophen) administration in subjects who consume moderate amounts of alcohol. Aliment Pharmacol Ther 2007;26(2):283–90.

172. Mofrad P, Contos MJ, Haque M, et al. Clinical and histologic spectrum of nonalcoholic fatty liver disease associated with normal ALT values. Hepatology 2003;37(6):1286–92.

173. Sanyal D, Mukherjee P, Raychaudhuri M, Ghosh S, Mukherjee S, Chowdhury S. Profile of liver enzymes in non-alcoholic fatty liver disease in patients with impaired glucose tolerance and newly detected untreated type 2 diabetes. Indian J Endocrinol Metab 2015;19(5):597–601.

174. Kim WR, Flamm SL, Bisceglie AM Di, Bodenheimer HC. SPECIAL ARTICLE Serum Activity of Alanine Aminotransferase (ALT) as an Indicator of Health and Disease. :1363–70.

175. Lominadze Z. Misconception : You Can ' t Have Liver Disease With Normal Liver Chemistries. 2018;12(4):96–9.

176. Elinav E, Ben-Dov IZ, Ackerman E, et al. Correlation between serum alanine aminotransferase activity and age: an inverted U curve pattern. Am J Gastroenterol 2005;100(10):2201–4.

177. Kotronen A, Juurinen L, Hakkarainen A, et al. Liver Fat Is Increased in Type 2 Diabetic Patients and Underestimated by Serum Alanine Aminotransferase Compared With Equally Obese Nondiabetic Subjects. Diabetes Care [Internet] 2008;31(1):165 LP – 169. Available from: http://care.diabetesjournals.org/content/31/1/165.abstract

178. Kaplan MM. Alanine aminotransferase levels: What's normal? Ann. Intern. Med. 2002;137(1):49.

179. Kunde SS, Lazenby AJ, Clements RH, Abrams GA. Spectrum of NAFLD and diagnostic implications of the proposed new normal range for serum ALT in obese women. Hepatology 2005;42(3):650–6.

180. Prati D, Taioli E, Zanella A, et al. Updated definitions of healthy ranges for serum alanine aminotransferase levels. Ann Intern Med 2002;137(1):1–10.

181. Reddy DC, Purushotham KR, Naidu BP. Circadian rhythmicity in aminotransferases in the Indian field mouse Mus booduga Gray. Indian J Exp Biol 1980;18(9):1049–50.

182. Dufour DR, Lott JA, Nolte FS, Gretch DR, Koff RS, Seeff LB. Diagnosis and monitoring of hepatic injury. I. Performance characteristics of laboratory tests. Clin Chem 2000;46(12):2027–49.

183. Ruhl CE, Everhart JE. Diurnal variation in serum alanine aminotransferase activity in the US population. J Clin Gastroenterol 2013;47(2):165–73.

184. Sattar N, McConnachie A, Ford I, et al. Serial Metabolic Measurements and Conversion to Type 2 Diabetes in the West of Scotland Coronary Prevention Study. Diabetes [Internet] 2007;56(4):984 LP – 991. Available from: http://diabetes.diabetesjournals.org/content/56/4/984.abstract

185. Taylor R. Pathogenesis of type 2 diabetes: tracing the reverse route from cure to cause. Diabetologia 2008;51(10):1781–9.

186. Waikar SS, Liu KD, Chertow GM. Diagnosis, epidemiology and outcomes of acute kidney injury. Clin J Am Soc Nephrol 2008;3(3):844–61.

187. Hébert HL, Shepherd B, Milburn K, et al. Cohort profile: Genetics of Diabetes Audit and Research in Tayside Scotland (GoDARTS). Int J Epidemiol 2018;47(2):380-381j.

188. McKinstry B, Sullivan FM, Vasishta S, et al. Cohort profile: The Scottish Research register SHARE. A register of people interested in research participation linked to NHS data sets. BMJ Open 2017;7(2):1–8.

189. Marcuccilli M, Chonchol M. NAFLD and chronic kidney disease. Int J Mol Sci 2016;17(4):1–15.

190. Romeo S, Kozlitina J, Xing C, et al. Genetic variation in PNPLA3 confers

susceptibility to nonalcoholic fatty liver disease. Nat Genet 2008;40(12):1461–5.

191. Wang X, Liu Z, Wang K, et al. Additive effects of the risk alleles of PNPLA3 and TM6SF2 on non-alcoholic fatty liver disease (NAFLD) in a Chinese population. Front Genet 2016;7(AUG):1–7.

192. Neuschwander-Tetri BA, Ünalp A, Creer MH. The upper limits of normal for serum ALT levels reported by clinical laboratories depend on local reference populations. Arch Intern Med 2004;168(6):663–6.

193. Younossi ZM. The Epidemiology of Nonalcoholic Steatohepatitis. Clinical Liver Disease. 2018;11(4):92–4.

194. Mantovani A, Scorletti E, Mosca A, Alisi A, Byrne CD, Targher G. Complications, morbidity and mortality of nonalcoholic fatty liver disease. Metabolism 2020;111S:154170.

195. Pradeepa R, Prabu AV, Jebarani S, Subhashini S, Mohan V. Use of a large diabetes electronic medical record system in India: clinical and research applications. 2011;

196. Mohammadi A, Bazazi A, Maleki-miyandoab T, Ghasemi-rad M. Evaluation of relationship between Grading of fatty liver and severity of Atherosclerotic finding. 2012;5(3):251–6.

197. Dasarathy S, Dasarathy J, Khiyami A, Joseph R, Lopez R, McCullough AJ. Validity of real time ultrasound in the diagnosis of hepatic steatosis: a prospective study. J Hepatol 2009;51(6):1061–7.

198. Ballestri S, Nascimbeni F, Lugari S, Lonardo A, Francica G. A critical appraisal of the use of ultrasound in hepatic steatosis. Expert Rev Gastroenterol Hepatol [Internet] 2019;13(7):667–81. Available from: https://doi.org/10.1080/17474124.2019.1621164

199. Joseph AE, Saverymuttu SH, al-Sam S, Cook MG, Maxwell JD. Comparison of liver histology with ultrasonography in assessing diffuse parenchymal liver disease. Clin Radiol 1991;43(1):26–31.

200. Younossi Z, Loomba R, Rinella M, Bugianesi E, Marchesini B. Current and Future Therapeutic Regimens for Non-alcoholic Fatty Liver Disease (NAFLD) and Non-alcoholic Steatohepatitis (NASH). Hepatology 2017;(5):1–36.

201. Michelotti GA, Machado M V, Diehl AM. NAFLD, NASH and liver cancer. Nat Rev Gastroenterol &Amp; Hepatol [Internet] 2013;10:656. Available from: https://doi.org/10.1038/nrgastro.2013.183

202. Kim GA, Lee HC, Choe J, et al. Association between non-alcoholic fatty liver disease and cancer incidence rate. J Hepatol [Internet] 2018;68(1):140–6. Available from: https://doi.org/10.1016/j.jhep.2017.09.012

203. Kanwal F, Kramer JR, Mapakshi S, et al. Risk of hepatocellular cancer in patients with non-alcoholic fatty liver disease. Gastroenterology 2018;155(6):1828–37.

204. White DL, Kanwal F, El-Serag HB. Association between nonalcoholic fatty liver disease and risk for hepatocellular cancer, based on systematic review. Clin Gastroenterol Hepatol 2012;10(12):1342-1359.e2.

205. Dongiovanni P, Romeo S, Valenti L. Hepatocellular carcinoma in nonalcoholic fatty liver: Role of environmental and genetic factors. World J Gastroenterol 2014;20(36):12945–55.

206. Wu J. Utilization of animal models to investigate nonalcoholic steatohepatitis-associated hepatocellular carcinoma. Oncotarget 2016;7(27):42762.

207. Yoon SK. Molecular mechanism of hepatocellular carcinoma. Hepatoma Res

2018;4(8):42.

208. Hwang ST, Cho YK, Park JH, et al. Relationship of non-alcoholic fatty liver disease to colorectal adenomatous polyps. J Gastroenterol Hepatol 2010;25(3):562–7.

209. Lee YI, Lim Y, Park HS. Colorectal neoplasms in relation to non-alcoholic fatty liver disease in Korean women: a retrospective cohort study. J Gastroenterol Hepatol 2012;27(1):91–5.

210. Wong VW-S, Wong GL-H, Tsang SW-C, et al. High prevalence of colorectal neoplasm in patients with non-alcoholic steatohepatitis. Gut 2011;60(6):829–36.

211. Stadlmayr A, Aigner E, Steger B, et al. Nonalcoholic fatty liver disease: an independent risk factor for colorectal neoplasia. J Intern Med 2011;270(1):41–9.

212. Huang K, Leu H, Wang Y, et al. Patients with nonalcoholic fatty liver disease have higher risk of colorectal adenoma after negative baseline colonoscopy. Color Dis 2013;15(7):830–5.

213. Sorensen HT, Mellemkjaer L, Jepsen P, et al. Risk of cancer in patients hospitalized with fatty liver: a Danish cohort study. J Clin Gastroenterol 2003;36(4):356–9.

214. Sanna C, Rosso C, Marietti M, Bugianesi E. Non-Alcoholic Fatty Liver Disease and Extra-Hepatic Cancers. 2016;

215. Chang C-F, Tseng Y-C, Huang H-H, Shih Y-L, Hsieh T-Y, Lin H-H. Exploring the relationship between nonalcoholic fatty liver disease and pancreatic cancer by computed tomographic survey. Intern Emerg Med 2018;13(2):191–7.

216. Kwak M-S, Yim JY, Yi A, et al. Nonalcoholic fatty liver disease is associated with breast cancer in nonobese women. Dig Liver Dis 2019;51(7):1030–5.

217. Wang Z, Zhao X, Chen S, et al. Associations Between Nonalcoholic Fatty Liver Disease and Cancers in a Large Cohort in China. Clin Gastroenterol Hepatol 2020;

218. Zhu C-Y, Qu J-C, Cao H-X, Chen G-Y, Shi Y-H, Fan J-G. Obesity and nonalcoholic fatty liver disease associated with adenocarcinoma in patients with lung cancer. Medicine (Baltimore) 2019;98(37):e17098.

219. Gaggini M, Morelli M, Buzzigoli E, DeFronzo RA, Bugianesi E, Gastaldelli A. Non-alcoholic fatty liver disease (NAFLD) and its connection with insulin resistance, dyslipidemia, atherosclerosis and coronary heart disease. Nutrients 2013;5(5):1544–60.

220. Brooke Steele C, Thomas CC, Jane Henley S, et al. Vital signs: Trends in incidence of cancers associated with overweight and obesity — United States, 2005-2014. Morb Mortal Wkly Rep 2017;66(39):1052–8.

221. Wolin KY, Carson K, Colditz GA. Obesity and cancer. Oncologist 2010;15(6):556–65.

222. Secretan BL, Ph D, Scoccianti C, Ph D, Loomis D, Ph D. Spe ci a l R e p or t Body Fatness and Cancer — Viewpoint of the IARC Working Group. 2019;

223. Polyzos SA, Kountouras J, Mantzoros CS. Adipokines in nonalcoholic fatty liver disease. Metabolism 2016;65(8):1062–79.

224. Morris AD, Boyle DI, MacAlpine R, et al. The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record linkage to create a diabetes register. DARTS/MEMO Collaboration. BMJ 1997;315(7107):524–8.

225. WHO | International Classification of Diseases, 11th Revision (ICD-11). WHO 2019;

226. Ioannou GN, Weiss NS, Boyko EJ, Mozaffarian D, Lee SP. Elevated serum alanine aminotransferase activity and calculated risk of coronary heart disease in the United States. Hepatology 2006;43(5):1145–51.

227. Hickman IJ, Russell AJ, Prins JB, Macdonald GA. Should patients with type 2 diabetes and raised liver enzymes be referred for further evaluation of liver disease? Diabetes Res Clin Pract 2008;80(1):2007–9.

228. Andersson HI, Ejlertsson G, Leden I, Rosenberg C. Chronic pain in a geographically defined general population: studies of differences in age, gender, social class, and pain localization. Clin J Pain [Internet] 1993;9(3):174—182. Available from: http://europepmc.org/abstract/MED/8219517

229. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med 2008;27(8):1133–63.

230. Weisell RC. Body mass index as an indicator of obesity. 2002;11.

231. Fleming M, Kirby B, Penny KI. Record linkage in Scotland and its applications to health research. J Clin Nurs 2012;21(19–20):2711–21.

232. Liu YL, Patman GL, Leathart JBS, et al. Carriage of the PNPLA3 rs738409 C >g polymorphism confers an increased risk of non-alcoholic fatty liver disease associated hepatocellular carcinoma. J Hepatol [Internet] 2014;61(1):75–81. Available from: http://dx.doi.org/10.1016/j.jhep.2014.02.030

233. Giovannucci E. Metabolic syndrome, hyperinsulinemia, and colon cancer: A review. Am J Clin Nutr 2007;86(3):836–42.

234. Tsujimoto T, Kajio H, Sugiyama T. Association between hyperinsulinemia and increased risk of cancer death in nonobese and obese people: A population-based observational study. Int J cancer 2017;141(1):102–11.

235. Kotronen A, Westerbacka J, Bergholm R, Pietiläinen KH, Yki-Järvinen H. Liver fat in

the metabolic syndrome. J Clin Endocrinol Metab 2007;92(9):3490–7.

236. Draznin B. Mechanism of the mitogenic influence of hyperinsulinemia. Diabetol Metab Syndr 2011;3(1):2–4.

237. Gao B, Tsukamoto H. Inflammation in Alcoholic and Nonalcoholic Fatty Liver Disease: Friend or Foe? Gastroenterology [Internet] 2016;150(8):1704–9. Available from: https://pubmed.ncbi.nlm.nih.gov/26826669

238. Shacter E, Weitzman SA. Chronic inflammation and cancer. 2002;

239. JARRAR MH, BARANOVA A, COLLANTES R, et al. Adipokines and cytokines in non-alcoholic fatty liver disease. Aliment Pharmacol Ther [Internet] 2008;27(5):412–21. Available from: https://doi.org/10.1111/j.1365-2036.2007.03586.x

240. Meissner HI, Breen N, Klabunde CN, Vernon SW. Patterns of colorectal cancer screening uptake among men and women in the United States. Cancer Epidemiol Prev Biomarkers 2006;15(2):389–94.

241. Maheswaran R, Pearson T, Jordan H, Black D. Socioeconomic deprivation, travel distance, location of service, and uptake of breast cancer screening in North Derbyshire, UK. J Epidemiol Community Heal 2006;60(3):208–12.

242. Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. Nat Rev Cancer 2005;5(11):845–56.

243. Adab P, Pallan M, Whincup PH. Is BMI the best measure of obesity ? 2018;1274(March):1–2. Available from: http://dx.doi.org/doi:10.1136/bmj.k1274

244. Williams CD, Stengel J, Asike MI, et al. Prevalence of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis among a largely middle-aged population utilizing ultrasound and liver biopsy: a prospective study. Gastroenterology 2011;140(1):124–

31.

245. Kahali B, Halligan B, Speliotes EK. Insights from Genome-Wide Association Analyses of Nonalcoholic Fatty Liver Disease. Semin Liver Dis 2015;35(4):375–91.

246. Collette L, Bogaerts J, Suciu S, et al. Statistical methodology for personalized medicine: New developments at EORTC Headquarters since the turn of the 21 st Century. Eur J Cancer, Suppl 2012;10(1):13–9.

247. Severson TJ, Besur S, Bonkovsky HL. Genetic factors that affect nonalcoholic fatty liver disease: A systematic clinical review. World J Gastroenterol 2016;22(29):6742–56.

248. Pirazzi C, Adiels M, Burza MA, et al. Patatin-like phospholipase domain-containing 3 (PNPLA3) I148M (rs738409) affects hepatic VLDL secretion in humans and in vitro. J Hepatol 2012;57(6):1276–82.

249. Romeo S, Sentinelli F, Dash S, et al. Morbid obesity exposes the association between PNPLA3 I148M (rs738409) and indices of hepatic injury in individuals of European descent. Int J Obes 2010;34(1):190–4.

250. Chung GE, Lee Y, Yim JY, et al. Genetic Polymorphisms of PNPLA3 and SAMM50 Are Associated with Nonalcoholic Fatty Liver Disease in a Korean Population. Gut Liver [Internet] 2017;1–8. Available from: http://www.gutnliver.org/journal/view.html?doi=10.5009/gnl17306

251. Kozlitina J, Smagris E, Stender S, et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. Nat Genet 2014;46(4):352–6.

252. Kitamoto T, Kitamoto A, Yoneda M, et al. Genome-wide scan revealed that

polymorphisms in the PNPLA3, SAMM50, and PARVB genes are associated with development and progression of nonalcoholic fatty liver disease in Japan. Hum Genet 2013;132(7):783–92.

253. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 2019;47(D1):D1005–12.

254. Bhatt SP, Misra A, Pandey RM. rs7903146 (C/T) polymorphism of Transcription factor 7 like 2 (TCF7L-2) gene is independently associated with non-alcoholic fatty liver disease in Asian Indians. Diabetes Metab Syndr Clin Res Rev [Internet] 2020;14(3):175–80. Available from: http://www.sciencedirect.com/science/article/pii/S1871402120300321

255. Bhatt SP, Nigam P, Misra A, Guleria R, Pandey RM, Pasha MAQ. Genetic variation in the patatin-like phospholipase domain-containing protein-3 (PNPLA-3) gene in Asian Indians with nonalcoholic fatty liver disease. Metab Syndr Relat Disord 2013;11(5):329–35.

256. Bale G, Steffie AU, Ravi Kanth VV, et al. Regional differences in genetic susceptibility to non-alcoholic liver disease in two distinct Indian ethnicities. World J Hepatol [Internet] 2017;9(26):1101–7. Available from: https://pubmed.ncbi.nlm.nih.gov/28989566

257. Wells JCK, Pomeroy E, Walimbe SR, Popkin BM, Yajnik CS. The Elevated Susceptibility to Diabetes in India: An Evolutionary Perspective. Front public Heal [Internet] 2016;4:145. Available from: https://pubmed.ncbi.nlm.nih.gov/27458578

258. Holliday EG. Hints of unique genetic effects for type 2 diabetes in India. Diabetes [Internet] 2013;62(5):1369–70. Available from:

https://pubmed.ncbi.nlm.nih.gov/23613552

259. Chambers JC, Eda S, Bassett P, et al. C-reactive protein, insulin resistance, central obesity, and coronary heart disease risk in Indian Asians from the United Kingdom compared with European whites. Circulation 2001;104(2):145–50.

260. Duseja A. Nonalcoholic fatty liver disease in India--is it different? Trop Gastroenterol Off J Dig Dis Found 2006;27(4):142–6.

261. Petersen KF, Dufour S, Feng J, et al. Increased prevalence of insulin resistance and nonalcoholic fatty liver disease in Asian-Indian men. Proc Natl Acad Sci [Internet] 2006;103(48):18273 LP – 18277. Available from: http://www.pnas.org/content/103/48/18273.abstract

262. Marees AT, de Kluiver H, Stringer S, et al. A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. Int J Methods Psychiatr Res 2018;27(2):e1608.

263. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81(3):559–75.

264. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protoc 2010;5(9):1564–73.

265. Laurie CC, Doheny KF, Mirel DB, et al. Quality control and quality assurance in genotypic data for genome-wide association studies. Genet Epidemiol 2010;34(6):591–602.

266. Joiret M, Mahachie John JM, Gusareva ES, Van Steen K. Confounding of linkage

disequilibrium patterns in large scale DNA based gene-gene interaction studies. BioData Min [Internet] 2019;12(1):11. Available from: https://doi.org/10.1186/s13040-019-0199-7

267. Astle W, Balding DJ. Population structure and cryptic relatedness in genetic association studies. Stat Sci 2009;24(4):451–71.

268. Consortium 1000 Genomes Project. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491(7422):56–65.

269. Graffelman J, van Eeuwijk F. Calibration of multivariate scatter plots for exploratory analysis of relations within and between sets of variables in genomic research. Biom J [Internet] 2005;47(6):863–79. Available from: http://europepmc.org/abstract/MED/16450858

270. Turner SD. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. Biorxiv 2014;5165.

271. Team RC. R: A language and environment for statistical computing. 2013;

272. Lin J-P, O'Donnell CJ, Fox CS, Cupples LA. Heritability of serum gamma-glutamyltransferase level: genetic analysis from the Framingham Offspring Study. Liver Int. Off. J. Int. Assoc. Study Liver. 2009;29(5):776–7.

273. Chambers JC, Zhang W, Sehmi J, et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. Nat Genet 2011;43(11):1131–8.

274. Zhang L, You W, Zhang H, et al. PNPLA3 polymorphisms (rs738409) and non-alcoholic fatty liver disease risk and related phenotypes: a meta-analysis. J Gastroenterol Hepatol [Internet] 2015;30(5):821–9. Available from:

https://doi.org/10.1111/jgh.12889

275. Dai G, Liu P, Li X, Zhou X, He S. Association between PNPLA3 rs738409 polymorphism and nonalcoholic fatty liver disease (NAFLD) susceptibility and severity: A meta-analysis. Medicine (Baltimore) 2019;98(7):e14324.

276. Singal AG, Manjunath H, Yopp AC, et al. The effect of PNPLA3 on fibrosis progression and development of hepatocellular carcinoma: a meta-analysis. Am J Gastroenterol 2014;109(3):325.

277. Trépo E, Romeo S, Zucman-Rossi J, Nahon P. PNPLA3 gene in liver diseases. J Hepatol 2016;65(2):399–412.

278. Pingitore P, Dongiovanni P, Motta BM, et al. PNPLA3 overexpression results in reduction of proteins predisposing to fibrosis. Hum Mol Genet [Internet] 2016;25(23):5212–22. Available from: https://doi.org/10.1093/hmg/ddw341

279. He S, McPhaul C, Li JZ, et al. A sequence variation (I148M) in PNPLA3 associated with nonalcoholic fatty liver disease disrupts triglyceride hydrolysis. J Biol Chem 2010;285(9):6706–15.

280. Rotman Y, Koh C, Zmuda JM, Kleiner DE, Liang TJ. The Association of Genetic Variability in Patatin-Like Phospholipase Domain-Containing Protein 3 (PNPLA3) with Histological Severity of Nonalcoholic Fatty Liver Disease. 2010;3:894–903.

281. Bruschi FV, Tardelli M, Claudel T, Trauner M. PNPLA3 expression and its impact on the liver: current perspectives. Hepat Med [Internet] 2017;9:55–66. Available from: https://pubmed.ncbi.nlm.nih.gov/29158695

282. Stickel F, Hampe J, Trépo E, Datz C, Romeo S. PNPLA3 genetic variation in alcoholic steatosis and liver disease progression. Hepatobiliary Surg Nutr [Internet]

2015;4(3):152–60. Available from: https://pubmed.ncbi.nlm.nih.gov/26151055

283. Pirazzi C, Valenti L, Motta BM, et al. PNPLA3 has retinyl-palmitate lipase activity in human hepatic stellate cells. Hum Mol Genet 2014;23(15):4077–85.

284. Bruschi FV, Claudel T, Tardelli M, et al. The PNPLA3 I148M variant modulates the fibrogenic phenotype of human hepatic stellate cells. Hepatology 2017;65(6):1875–90.

285. Palmer ND, Musani SK, Yerges-Armstrong LM, et al. Characterization of European ancestry nonalcoholic fatty liver disease-associated variants in individuals of African and Hispanic descent. Hepatology 2013;58(3):966–75.

286. Torres DM, Harrison SA. NAFLD: Predictive value of ALT levels for NASH and advanced fibrosis. Nat Rev Gastroenterol Hepatol [Internet] 2013;10(9):510–1. Available from: http://dx.doi.org/10.1038/nrgastro.2013.138

287. Yuan X, Waterworth D, Perry JRB, et al. Population-Based Genome-wide Association Studies Reveal Six Loci Influencing Plasma Levels of Liver Enzymes. Am J Hum Genet 2008;83(4):520–8.

288. Simons K, Ikonen E. Functional rafts in cell membranes. Nature 1997;387(6633):569–72.

289. Zhang X-Q, Xu C-F, Yu C-H, Chen W-X, Li Y-M. Role of endoplasmic reticulum stress in the pathogenesis of nonalcoholic fatty liver disease. World J Gastroenterol [Internet] 2014;20(7):1768–76. Available from: https://pubmed.ncbi.nlm.nih.gov/24587654

290. Huber MD, Vesely PW, Datta K, Gerace L. Erlins restrict SREBP activation in the ER and regulate cellular cholesterol homeostasis. J Cell Biol [Internet] 2013;203(3):427–36. Available from: https://doi.org/10.1083/jcb.201305076

291. Wang X, Sato R, Brown MS, Hua X, Goldstein JL. SREBP-1, a membrane-bound transcription factor released by sterol-regulated proteolysis. Cell 1994;77(1):53–62.

292. Brown MS, Goldstein JL. The SREBP Pathway : Regulation of Cholesterol Metabolism by Proteolysis of a Membrane-Bound Transcription Factor. Cell. 1997;89(1):331–40.

293. Shimano H, Sato R. SREBP-regulated lipid metabolism: convergent physiology — divergent pathophysiology. Nat Rev Endocrinol [Internet] 2017;13(12):710–30. Available from: https://doi.org/10.1038/nrendo.2017.91

294. Moslehi A, Hamidi-Zad Z. Role of SREBPs in Liver Diseases: A Mini-review. J Clin Transl Hepatol [Internet] 2018;6(3):332–8. Available from: https://pubmed.ncbi.nlm.nih.gov/30271747

295. Ponugoti B, Fang S, Kemper JK. Functional Interaction of Hepatic Nuclear Factor-4 and Peroxisome Proliferator-Activated Receptor-γ Coactivator 1α in CYP7A1 Regulation Is Inhibited by a Key Lipogenic Activator, Sterol Regulatory Element-Binding Protein-1c. Mol Endocrinol [Internet] 2007;21(11):2698–712. Available from: https://doi.org/10.1210/me.2007-0196

296. Feitosa MF, Wojczynski MK, North KE, et al. The ERLIN1-CHUK-CWF19L1 gene cluster influences liver fat deposition and hepatic inflammation in the NHLBI Family Heart Study. Atherosclerosis 2013;228(1):175–80.

297. Rotman Y, Koh C, Zmuda JM, Kleiner DE, Liang TJ, CRN N. The association of genetic variability in patatin-like phospholipase domain-containing protein 3 (PNPLA3) with histological severity of nonalcoholic fatty liver disease. Hepatology 2010;52(3):894–903.

298. Barkett M, Gilmore TD. Control of apoptosis by Rel/NF-kappaB transcription factors. Oncogene 1999;18(49):6910–24.

299. Jansen R, Hottenga J-J, Nivard MG, et al. Conditional eQTL analysis reveals allelic heterogeneity of gene expression. Hum Mol Genet 2017;26(8):1444–51.

300. Hewitt J, Walters M, Padmanabhan S, Dawson J. Cohort profile of the UK Biobank: diagnosis and characteristics of cerebrovascular disease. BMJ Open [Internet] 2016;6(3):e009161. Available from: http://bmjopen.bmj.com/content/6/3/e009161.abstract

301. Uhlén M, Fagerberg L, Hallström BM, et al. Proteomics. Tissue-based map of the human proteome. Science 2015;347(6220):1260419.

302. Hoffmann TJ, Theusch E, Haldar T, et al. A large electronic-health-record-based genome-wide study of serum lipids. Nat Genet 2018;50(3):401–13.

303. Wood AR, Perry JRB, Tanaka T, et al. Imputation of variants from the 1000 Genomes Project modestly improves known associations and can identify low-frequency variant-phenotype associations undetected by HapMap based imputation. PLoS One [Internet] 2013;8(5):e64343–e64343. Available from: https://pubmed.ncbi.nlm.nih.gov/23696881

304. Schwabe GC, Hoffmann K, Loges NT, et al. Primary ciliary dyskinesia associated with normal axoneme ultrastructure is caused by DNAH11 mutations. Hum Mutat 2008;29(2):289–98.

305. Wang W, Li T, Wang X, et al. FAM19A4 is a novel cytokine ligand of formyl peptide receptor 1 (FPR1) and is able to promote the migration and phagocytosis of macrophages. Cell Mol Immunol 2015;12(5):615–24.

306. Võsa U, Claringbould A, Westra H-J, et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. bioRxiv [Internet] 2018;447367. Available from: http://biorxiv.org/content/early/2018/10/19/447367.abstract

307. Sakaidani Y, Nomura T, Matsuura A, et al. O-Linked-N-acetylglucosamine on extracellular protein domains mediates epithelial cell–matrix interactions. Nat Commun [Internet] 2011;2(1):583. Available from: https://doi.org/10.1038/ncomms1591

308. Zhang K, Yin R, Yang X. O-GlcNAc: A Bittersweet Switch in Liver. Front Endocrinol (Lausanne) 2014;5:221.

309. Dentin R, Liu Y, Koo S-H, et al. Insulin modulates gluconeogenesis by inhibition of the coactivator TORC2. Nature 2007;449(7160):366–9.

310. Yang X, Ongusaha PP, Miles PD, et al. Phosphoinositide signalling links O-GlcNAc transferase to insulin resistance. Nature 2008;451(7181):964–9.

311. Elsworth B, Lyon M, Alexander T, et al. The MRC IEU OpenGWAS data infrastructure. bioRxiv [Internet] 2020;2020.08.10.244293. Available from: http://biorxiv.org/content/early/2020/08/10/2020.08.10.244293.abstract

312. Hoffmann TJ, Ehret GB, Nandakumar P, et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. Nat Genet 2017;49(1):54–64.

313. Facchinello N, Tarifeño-Saldivia E, Grisan E, et al. Tcf7l2 plays pleiotropic roles in the control of glucose homeostasis, pancreas morphology, vascularization and regeneration. Sci Rep [Internet] 2017;7(1):9605. Available from: https://pubmed.ncbi.nlm.nih.gov/28851992

314. Imamura M, Takahashi A, Yamauchi T, et al. Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes. Nat Commun 2016;7:10531.

315. Ghassibe-Sabbagh M, Haber M, Salloum AK, et al. T2DM GWAS in the Lebanese population confirms the role of TCF7L2 and CDKAL1 in disease susceptibility. Sci Rep 2014;4(1):1–9.

316. Sousa AGP, Marquezine GF, Lemos PA, et al. TCF7L2 polymorphism rs7903146 is associated with coronary artery disease severity and mortality. PLoS One 2009;4(11):e7697.

317. Villareal DT, Robertson H, Bell GI, et al. TCF7L2 variant rs7903146 affects the risk of type 2 diabetes by modulating incretin action. Diabetes 2010;59(2):479–85.

318. Musso G, Gambino R, Pacini G, Pagano G, Durazzo M, Cassader M. Transcription factor 7-like 2 polymorphism modulates glucose and lipid homeostasis, adipokine profile, and hepatocyte apoptosis in NASH. Hepatology 2009;49(2):426–35.

319. Zain SM, Mohamed R, Mahadeva S, et al. A multi-ethnic study of a PNPLA3 gene variant and its association with disease severity in non-alcoholic fatty liver disease. Hum Genet 2012;131(7):1145–52.

320. Pesce M, Schöler HR. Oct-4: gatekeeper in the beginnings of mammalian development. Stem Cells 2001;19(4):271–8.

321. Chien Y, Huang C-S, Lin H-C, et al. Improvement of non-alcoholic steatohepatitis by hepatocyte-like cells generated from iPSCs with Oct4/Sox2/Klf4/Parp1. Oncotarget [Internet] 2018;9(26):18594–606. Available from: https://pubmed.ncbi.nlm.nih.gov/29719629

322. Park MR, Wong MS, Arauzo-Bravo MJ, et al. Oct4 and Hnf4α-induced hepatic stem cells ameliorate chronic liver injury in liver fibrosis model. PLoS One 2019;14(8):e0221085.

323. Zhu P, Wang Y, He L, et al. ZIC2-dependent OCT4 activation drives self-renewal of human liver cancer stem cells. J Clin Invest 2015;125(10):3795–808.

324. Han P, Werber J, Surana M, Fleischer N, Michaeli T. The calcium/calmodulin-dependent phosphodiesterase PDE1C down-regulates glucose-induced insulin secretion. J Biol Chem 1999;274(32):22337–44.

325. Junker AE, Gluud L, Holst JJ, Knop FK, Vilsbøll T. Diabetic and nondiabetic patients with nonalcoholic fatty liver disease have an impaired incretin effect and fasting hyperglucagonaemia. J Intern Med [Internet] 2016;279(5):485–93. Available from: https://doi.org/10.1111/joim.12462

326. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature [Internet] 2020;581(7809):434–43. Available from: https://doi.org/10.1038/s41586-020-2308-7

327. Narayanasamy K, Karthick R, Panneerselvam P, et al. Association of metabolic syndrome and patatin-like phospholipase 3–rs738409 gene variant in non-alcoholic fatty liver disease among a Chennai-based south Indian population. J Gene Med 2020;22(4):e3160.

328. Thorlund K, Imberger G, Johnston BC, et al. Evolution of heterogeneity (I 2) estimates and their 95% confidence intervals in large meta-analyses. PLoS One 2012;7(7):e39471.

329. Lipsey MW, Aiken LS. Design sensitivity: Statistical power for experimental research.

sage; 1990.

330. Wu M-J, Yuan C, Lu L-L, An B-Q, Xuan S-Y, Xin Y-N. Role of NCAN rs2228603 polymorphism in the incidence of nonalcoholic fatty liver disease: a case-control study. Lipids Health Dis [Internet] 2016;15(1):207. Available from: https://pubmed.ncbi.nlm.nih.gov/27887608

331. Cohen J. Statistical power analysis. Curr Dir Psychol Sci 1992;1(3):98–101.

332. Chalasani N, Guo X, Loomba R, et al. Genome-wide association study identifies variants associated with histologic features of nonalcoholic fatty liver disease. Gastroenterology 2010;139(5):1567–76.

333. Di Costanzo A, Belardinilli F, Bailetti D, et al. Evaluation of Polygenic Determinants of Non-Alcoholic Fatty Liver Disease (NAFLD) By a Candidate Genes Resequencing Strategy. Sci Rep [Internet] 2018;8(1):3702. Available from: https://pubmed.ncbi.nlm.nih.gov/29487372

334. Edelman D, Kalia H, Delio M, et al. Genetic analysis of nonalcoholic fatty liver disease within a Caribbean-Hispanic population. Mol Genet genomic Med [Internet] 2015;3(6):558–69. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26740948%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4694126

335. Gorden A, Yang R, Yerges-Armstrong LM, et al. Genetic variation at NCAN locus is associated with inflammation and fibrosis in non-alcoholic fatty liver disease in morbid obesity. Hum Hered 2013;75(1):34–43.

336. Hernaez R, McLean J, Lazo M, et al. Association between variants in or near PNPLA3, GCKR, and PPP1R3B with ultrasound-defined steatosis based on data from

the third National Health and Nutrition Examination Survey. Clin Gastroenterol Hepatol [Internet] 2013;11(9):1183-1190.e2. Available from: https://pubmed.ncbi.nlm.nih.gov/23416328

337. Mahdessian H, Taxiarchis A, Popov S, Silveira A, Franco-cereceda A. TM6SF2 is a regulator of liver fat metabolism influencing triglyceride secretion and hepatic lipid droplet content. Proceedings of the National Academy of Sciences. 2014;111(24).

338. Liu Y-L, Reeves HL, Burt AD, et al. TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. Nat Commun [Internet] 2014;5:1–6. Available from: http://www.nature.com/doifinder/10.1038/ncomms5309

339. Li Y, Liu S, Gao Y, et al. Association of TM6SF2 rs58542926 gene polymorphism with the risk of non-alcoholic fatty liver disease and colorectal adenoma in Chinese Han population. BMC Biochem [Internet] 2019;20(1):3. Available from: https://doi.org/10.1186/s12858-019-0106-3

340. Fracanzani AL, Valenti L, Bugianesi E, et al. Risk of severe liver disease in nonalcoholic fatty liver disease with normal aminotransferase levels: A role for insulin resistance and diabetes. Hepatology 2008;48(3):792–8.

341. Singh D, Das CJ, Baruah MP. Imaging of non alcoholic fatty liver disease: A road less travelled. Indian J Endocrinol Metab [Internet] 2013;17(6):990–5. Available from: https://pubmed.ncbi.nlm.nih.gov/24381873

342. Strauss S, Gavish E, Gottlieb P, Katsnelson L. Interobserver and intraobserver variability in the sonographic assessment of fatty liver. Am J Roentgenol 2007;189(6):W320–3.

343. Marchesini G, Bugianesi E, Forlani G, et al. Nonalcoholic fatty liver, steatohepatitis,

and the metabolic syndrome. Hepatology 2003;37(4):917–23.

344. Sánchez-Garrido MA, Brandt SJ, Clemmensen C, Müller TD, DiMarchi RD, Tschöp MH. GLP-1/glucagon receptor co-agonism for treatment of obesity. Diabetologia 2017;60(10):1851–61.

345. Baggio LL, Drucker DJ. Biology of Incretins : GLP-1 and GIP. 2007;2131–57.

346. Jiang G, Zhang BB. Glucagon and regulation of glucose metabolism. Am J Physiol Metab 2003;284(4):E671–8.

347. Patel V, Joharapurkar A, Dhanesha N, et al. Co-agonist of glucagon and GLP-1 reduces cholesterol and improves insulin sensitivity independent of its effect on appetite and body weight in diet-induced obese C57 mice. Can J Physiol Pharmacol [Internet] 2013;91(12):1009–15. Available from: https://doi.org/10.1139/cjpp-2013-0189

348. Hall E, Dayeh T, Kirkpatrick CL, Wollheim CB, Dekker Nitert M, Ling C. DNA methylation of the glucagon-like peptide 1 receptor (GLP1R) in human pancreatic islets. BMC Med Genet [Internet] 2013;14(1):76. Available from: https://doi.org/10.1186/1471-2350-14-76

349. Barbato A, Russo P, Venezia A, Strazzullo V, Siani A, Cappuccio FP. Analysis of Gly40Ser polymorphism of the glucagon receptor (GCGR) gene in different ethnic groups. J Hum Hypertens 2003;17(8):577.

350. Sathananthan A, Man CD, Micheletto F, et al. Common genetic variation in GLP1R and insulin secretion in response to exogenous GLP-1 in nondiabetic subjects: a pilot study. Diabetes Care 2010;33(9):2074–6.

351. Karczewski KJ, Francioli LC, Tiao G, et al. Variation across 141,456 human exomes

and genomes reveals the spectrum of loss-of- function intolerance across human protein-coding genes. Nature. 2019;

352. Sheikh HI, Dougherty LR, Hayden EP, Klein DN, Singh SM. Glucagon-like peptide-1 receptor gene polymorphism (Leu260Phe) is associated with morning cortisol in preschoolers. Prog Neuropsychopharmacol Biol Psychiatry 2010;34(6):980–3.

353. de Luis DA, Ballesteros M, Guzman AL, et al. rs6923761 gene variant in glucagon-like peptide 1 receptor: Allelic frequencies and influence on cardiovascular risk factors in a multicenter study of Castilla-Leon. Clin Nutr [Internet] 2018;37(6, Part A):2144–8. Available from: http://www.sciencedirect.com/science/article/pii/S0261561417313869

354. de Luis DA, Aller R, Izaola O, Bachiller R. Role of rs6923761 gene variant in glucagon-like peptide 1 receptor in basal GLP-1 levels, cardiovascular risk factor and serum adipokine levels in naïve type 2 diabetic patients. J Endocrinol Invest 2015;38(2):143–7.

355. Javorský M, Gotthardová I, Klimčáková L, et al. A missense variant in GLP1R gene is associated with the glycaemic response to treatment with gliptins. Diabetes, Obes Metab [Internet] 2016;18(9):941–4. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/dom.12682

356. Sandoval DA, Alessio DAD. PHYSIOLOGY OF PROGLUCAGON PEPTIDES : ROLE OF GLUCAGON AND GLP-1 IN HEALTH AND DISEASE. Physiological reviews. 2020;513–48.

357. Lee J, Hong S-W, Rhee E-J, Lee W-Y. GLP-1 Receptor Agonist and Non-Alcoholic Fatty Liver Disease. Diabetes Metab J 2012;36(4):262–7.

358. Ben-Shlomo S, Zvibel I, Shnell M, et al. Glucagon-like peptide-1 reduces hepatic lipogenesis via activation of AMP-activated protein kinase. J Hepatol 2011;54(6):1214–23.

359. Zhang L, Yang M, Ren H, et al. GLP-1 analogue prevents NAFLD in ApoE KO mice with diet and Acrp30 knockdown by inhibiting c-JNK. Liver Int 2013;33(5):794–804.

360. Baggio LL, Drucker DJ. Biology of incretins: GLP-1 and GIP. Gastroenterology 2007;132(6):2131–57.

361. Junker AE. The role of incretin hormones and glucagon in patients with liver disease. Faculty of Health and Medical Sciences, University of Copenhagen; 2015.

362. Calanna S, Christensen M, Holst JJ, et al. Secretion of glucose-dependent insulinotropic polypeptide in patients with type 2 diabetes: systematic review and meta-analysis of clinical studies. Diabetes Care 2013;36(10):3346–52.

363. Liu Y, Wei R, Hong T. Potential roles of glucagon-like peptide-1-based therapies in treating non-alcoholic fatty liver disease. 2014;20(27):9090–7.

364. Zheng T, Chen B, Yang L, et al. Association of plasma dipeptidyl peptidase-4 activity with non-alcoholic fatty liver disease in nondiabetic Chinese population. Metabolism 2017;73:125–34.

365. Lee J, Hong S, Rhee E, Lee W. GLP-1 Receptor Agonist and Non-Alcoholic Fatty Liver Disease. 2012;262–7.

366. Cuthbertson DJ, Irwin A, Gardner CJ, et al. Improved glycaemia correlates with liver fat reduction in obese, type 2 diabetes, patients given glucagon-like peptide-1 (GLP-1) receptor agonists. PLoS One 2012;7(12):e50117.

367. Kim S, Jeong J, Jung H-S, et al. Anti-inflammatory effect of glucagon like peptide-1

receptor agonist, exendin-4, through modulation of IB1/JIP1 expression and JNK signaling in stroke. Exp Neurobiol 2017;26(4):227–39.

368. Somm E, Montandon SA, Loizides-Mangold U, et al. The GLP-1R agonist liraglutide limits hepatic lipotoxicity and inflammatory response in mice fed a methionine-choline deficient diet. Transl Res 2020;

369. Kirk M. Habegger, Kristy M. Heppner, Nori Geary, Timothy J. Bartness RD, Tschöp and MH. The metabolic actions of glucagon revisited. Nat Rev Endocrinol 6, 689–697 (2010). https://doi.org/10.1038/nrendo.2010.187

370. Kazda CM, Ding Y, Kelly RP, et al. Evaluation of Efficacy and Safety of the Glucagon Receptor Antagonist LY2409021 in Patients With Type 2 Diabetes: 12- and 24-Week Phase 2 Studies. Diabetes Care 2016;39(7):1241–9.

371. Nason SR, Kim T, Antipenko JP, et al. Glucagon-Receptor Signaling Reverses Hepatic Steatosis Independent of Leptin Receptor Expression. Endocrinology 2019;

372. Janah L, Kjeldsen S, Galsgaard KD, et al. Glucagon Receptor Signaling and Glucagon Resistance. Int J Mol Sci 2019;20(13).

373. Seghieri M, Christensen AS, Andersen A, Solini A, Knop FK, Vilsbøll T. Future Perspectives on GLP-1 Receptor Agonists and GLP-1/glucagon Receptor Co-agonists in the Treatment of NAFLD. Front Endocrinol (Lausanne) 2018;9:649.

374. SCHULMAN JL, CARLETON JL, WHITNEY G, WHITEHORN JC. Effect of glucagon on food intake and body weight in man. J Appl Physiol 1957;11(3):419–21.

375. Grimsby J, Pardo V, Rondinone CM, Valdecantos MP, Valverde ÁM, Álvarez C. Beneficial effects of a dual acting GLP-1R/glucagon receptor co-agonist in the treatment of hepatic regeneration in NAFLD [Internet]. 2014;Available from:

http://hdl.handle.net/10261/125630

376. Elvert R, Bossart M, Herling AW, et al. Team Players or Opponents: Coadministration of Selective Glucagon and GLP-1 Receptor Agonists in Obese Diabetic Monkeys. Endocrinology [Internet] 2018;159(8):3105–19. Available from: https://doi.org/10.1210/en.2018-00399

377. Kannt, A, Madsen, AN, Kammermeier, C, et al. Incretin combination therapy for the treatment of non-alcoholic steatohepatitis. Diabetes Obes Metab. 2020; 22: 1328–1338.

378. Ghosh A, Nair R. Improved Clinical Outcomes With Dulaglutide as Add-on Medication to Oral Antidiabetic Drugs With or Without Insulin in Overweight Indian Patients With Type 2 Diabetes Mellitus: Retrospective Study in a Real-World Setting. Curr Diabetes Rev 2020;16(5):490–6.

379. De Silva S, Li W, Kemos P, et al. Non-invasive markers of liver fibrosis in fatty liver disease are unreliable in people of South Asian descent. Frontline Gastroenterol [Internet] 2018;9(2):115–21. Available from: https://pubmed.ncbi.nlm.nih.gov/29588839

380. Pati GK, Singh SP. Nonalcoholic Fatty Liver Disease in South Asia. Euroasian J hepato-gastroenterology [Internet] 2016;6(2):154–62. Available from: https://pubmed.ncbi.nlm.nih.gov/29201749

381. Mofrad P, Contos MJ, Haque M, et al. Clinical and histologic spectrum of nonalcoholic fatty liver disease associated with normal ALT values. Hepatology 2003;37(6):1286–92.

382. Marchesini G, Brizi M, Morselli-Labate AM, et al. Association of nonalcoholic fatty

liver disease with insulin resistance. Am J Med [Internet] 1999;107(5):450–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10569299

383. Davenport AP, Hyndman KA, Dhaun N, et al. Endothelin. Pharmacol Rev 2016;68(2):357–418.

384. Degertekin B, Ozenirler S, Elbeg S, Akyol G. The Serum Endothelın-1 Level in Steatosıs and NASH, and Its Relatıon wıth Severıty of Lıver Fıbrosıs. Digestive diseases and sciences. 2007 Oct 1;52(10):2622-8.

385. Takeshi S, Masashi Y, Tomoh M. Molecular characterization of endothelin receptors. Trends Pharmacol Sci 1992;13:103–8.

386. Shetty SS, Okada T, Webb RL, DelGrande D, Lappe RW. Functionally distinct endothelin B receptors in vascular endothelium and smooth muscle. Biochem Biophys Res Commun 1993;191(2):459–64.

387. Gandhi CR, Stephenson K, Olson MS. Endothelin, a potent peptide agonist in the liver. J Biol Chem 1990;265(29):17432–5.

388. Moreira RK. Hepatic stellate cells and liver fibrosis. Arch Pathol Lab Med 2007;131(11):1728–34.

389. Yokomori H, Oda M, Ogi M, Kamegaya Y, Tsukada N, Nakamura M. Enhanced expression of endothelin receptor subtypes in cirrhotic rat liver. 2001;(8):114–22.

390. Channick RN, Sitbon O, Barst RJ, Manes A, Rubin LJ. Endothelin receptor antagonists in pulmonary arterial hypertension. J Am Coll Cardiol 2004;43(12 Supplement):S62–7.

391. Rockey DC, Chung JJ. Endothelin antagonism in experimental hepatic fibrosis. Implications for endothelin in the pathogenesis of wound healing. J Clin Invest

1996;98(6):1381–8.

392. Popescu A. Preliminary results of the treatment with an endothelin receptor antagonist (ET-RA) in patients with non-alcoholic steatohepatitis (NASH) associating metabolic disorders. J Hepatol [Internet] 2003;38:199. Available from: https://doi.org/10.1016/S0168-8278(03)80066-1

393. Limongelli G, Maddaloni V, Dario C, Michele V, Calabro P. Analysis of endothelin-1 and endothelin-1 receptor A gene polymorphisms in patients with pulmonary arterial hypertension. 2012;425–30.

394. Beaudoin M, Gupta RM, Won H-H, et al. Myocardial infarction–associated SNP at 6p24 interferes with MEF2 binding and associates with PHACTR1 expression levels in human coronary arteries. Arterioscler Thromb Vasc Biol 2015;35(6):1472–9.

395. Benjafield A V, Katyk K, Morris BJ. Association of EDNRA, but not WNK4 or FKBP1B, polymorphisms with essential hypertension. Clin Genet 2003;64(5):433–8.

396. Malik R, Chauhan G, Traylor M, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. Nat Genet 2018;50(4):524–37.

397. Hartiala J, Schwartzman WS, Gabbay J, Ghazalpour A, Bennett BJ, Allayee H. The Genetic Architecture of Coronary Artery Disease: Current Knowledge and Future Opportunities. Curr Atheroscler Rep 2017;19(2):6.

398. Sánchez-Mejías A, Fernández RM, López-Alonso M, Antiñolo G, Borrego S. New roles of EDNRB and EDN3 in the pathogenesis of Hirschsprung disease. Genet Med 2010;12(1):39–43.

399. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic

variation. Nucleic Acids Res [Internet] 2001;29(1):308–11. Available from: https://pubmed.ncbi.nlm.nih.gov/11125122

400. Abdullah Z, Hackstein P, Trebicka J, et al. Microbial translocation in liver fibrosis induces chronic IFNAR signaling that directly affects innate immune responses to systemic bacterial infection. Z Gastroenterol 2015;53(01):A5_2.

401. Lapointe-shaw L, Georgie F, Carlone D, et al. Identifying cirrhosis , decompensated cirrhosis and hepatocellular carcinoma in health administrative data : A validation study. PLOS Medicine 2018;1–13.

402. Nanji AA, French SW, Freeman JB. Serum Alanine Aminotransferase to Aspartate Aminotransferase Ratio and Degree of Fatty Liver in Morbidly Obese Patients. Enzyme [Internet] 1986;36:266–9. Available from: https://www.karger.com/DOI/10.1159/000469304

403. Yu AS, Keeffe EB. Elevated AST or ALT to nonalcoholic fatty liver disease: accurate predictor of disease prevalence? Am J Gastroenterol 2003;98(5):955–6.

404. Bis JC, Kavousi M, Franceschini N, et al. Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque. Nat Genet 2011;43(10):940.

405. Hemerich D, van der Laan SW, Tragante V, et al. Impact of carotid atherosclerosis loci on cardiovascular events. Atherosclerosis 2015;243(2):466–8.

406. Wannamethee G, Ebrahim S, Gerald Shaper A. Gamma-glutamyltransferase: determinants and association with mortality from ischemic heart disease and all causes. Am J Epidemiol 1995;142(7):699–708.

407. Nikpay M, Goel A, Won H-H, et al. A comprehensive 1000 Genomes–based genome-

wide association meta-analysis of coronary artery disease. Nat Genet 2015;47(10):1121.

408. Hager J, Kamatani Y, Cazier J-B, et al. Genome-wide association study in a Lebanese cohort confirms PHACTR1 as a major determinant of coronary artery stenosis. PLoS One 2012;7(6):e38663.

409. Kuveljic J, Djuric T, Stankovic A, Koncar I, Alavantic D, Zivkovic M. PHACTR1 haplotypes are associated with carotid plaque presence and affect PHACTR1 mRNA expression in carotid plaque tissue. Gene 2019;710:273–8.

410. Tsuchiya Y, Suzuki S, Inaba K, et al. Impact of Endothelin-1 on Microcirculatory Disturbance after Partial Hepatectomy under Ischemia / Reperfusion in Thioacetamide-Induced Cirrhotic Rats. Journal of Surgical Research 2003;108:100–8.

411. Møller S, Gülberg V, Henriksen JH, Gerbes AL. Endothelin-1 and endothelin-3 in cirrhosis: relations to systemic and splanchnic haemodynamics. J Hepatol 1995;23(2):135–44.

412. Hynynen MM, Khalil RA. The vascular endothelin system in hypertension--recent patents and discoveries. Recent Pat Cardiovasc Drug Discov 2006;1(1):95–108.

413. Buob S, Johnston AN, Webster CRL. Portal hypertension: pathophysiology, diagnosis, and treatment. J Vet Intern Med 2011;25(2):169–86.

414. Hennenberg M, Trebicka J, Sauerbruch T, Heller J. Mechanisms of extrahepatic vasodilation in portal hypertension. Gut [Internet] 2008;57(9):1300 LP – 1314. Available from: http://gut.bmj.com/content/57/9/1300.abstract

415. Møller S, Bendtsen F. The pathophysiology of arterial vasodilatation and hyperdynamic circulation in cirrhosis. Liver Int 2018;38(4):570–80.

416. Mari R, Mun A, Poo JL, et al. Chronic Blockade of Endothelin Receptors in Cirrhotic Rats : Hepatic and Hemodynamic Effects. 1999;161–7.

417. Sanyal AJ, Bosch J, Blei A, Arroyo V. Portal hypertension and its complications. Gastroenterology 2008;134(6):1715–28.

418. Ahlborg G, Weitzberg E, Lundberg JM. Circulating endothelin-1 reduces splanchnic and renal blood flow and splanchnic glucose production in humans. J Appl Physiol 1995;79(1):141–5.

419. Lautt WW. Regulatory processes interacting to maintain hepatic blood flow constancy: vascular compliance, hepatic arterial buffer response, hepatorenal reflex, liver regeneration, escape from vasoconstriction. Hepatol Res 2007;37(11):891–903.

420. Breslow MJ, Tobin JR, Bredt DS, Ferris CD, Snyder SH, Traystman RJ. Nitric oxide as a regulator of adrenal blood flow. Am J Physiol Circ Physiol 1993;264(2):H464–9.

421. Ijaz S, Yang W, Winslet MC, Seifalian AM. Impairment of hepatic microcirculation in fatty liver. Microcirculation 2003;10(6):447–56.

422. Chu MJJ, Hickey AJR, Phillips ARJ, Bartlett ASJR. The Impact of Hepatic Steatosis on Hepatic Ischemia-Reperfusion Injury in Experimental Studies : A Systematic Review. Biomed Research International. 2013;2013.

423. Pasarín M, Abraldes JG, Liguori E, Kok B, La Mura V. Intrahepatic vascular changes in non-alcoholic fatty liver disease: Potential role of insulin-resistance and endothelial dysfunction. World J Gastroenterol 2017;23(37):6777–87.

424. Kowalczyk A, Kleniewska P, Goraca A. The Role of Endothelin-1 and Endothelin Receptor Antagonists in Inflammatory Response and Sepsis. Arch. Immunol. Ther. Exp. 2015;41–52.

425. Ruetten H, Thiemermann C. Endothelin-1 stimulates the biosynthesis of tumour necrosis factor in macrophages: ET-receptors, signal transduction and inhibition by dexamethasone. J Physiol Pharmacol an Off J Polish Physiol Soc 1997;48(4):675–88.

426. Ohkita M, Tawa M, Kitada K, Matsumura Y. Pathophysiological roles of endothelin receptors in cardiovascular diseases. J Pharmacol Sci 2012;12R01CR.

427. Boffa J-J, Tharaux P-L, Dussaule J-C, Chatziantoniou C. Regression of renal vascular fibrosis by endothelin receptor antagonism. Hypertension 2001;37(2):490–6.

428. Guo C, Wu J, Wu Y, Zhong M, Lu H. Effects of endothelin-1 on hepatic stellate cell proliferation , collagen synthesis and secretion , intracellular free calcium concentration. World J Gastroenterol. 2004;10(18):2697–700.

429. Pinzani M, Milani S, De Franco R, et al. Endothelin 1 is overexpressed in human cirrhotic liver and exerts multiple effects on activated hepatic stellate cells. Gastroenterology 1996;110(2):534–48.

430. Fox BM, Kasztan M. Endothelin receptor antagonists in sickle cell disease: A promising new therapeutic approach. Life Sci 2016;159:15–9.

431. Carducci MA, Jimeno A. Targeting bone metastasis in prostate cancer with endothelin receptor antagonists. Clin Cancer Res 2006;12(20):6296s-6300s.

432. Neuhofer W, Pittrow D. Role of endothelin and endothelin receptor antagonists in renal disease. Eur J Clin Invest 2006;36:78–88.

433. Cavasin MA, Semus H, Pitts K, et al. Acute effects of endothelin receptor antagonists on hepatic hemodynamics of cirrhotic and noncirrhotic rats 1. Canadian Journal of Physiology and Pharmacology 2010;643:636–43.

434. De Gottardi A, Shaw S, Sagesser H, Reichen J. Type A, but not type B, endothelin

receptor antagonists significantly decrease portal pressure in portal hypertensive rats. J Hepatol 2000;33(5):733–7.

435. Hoeper MM. Liver toxicity: the Achilles' heel of endothelin receptor antagonist therapy? Eur Respiratory Soc . 2009;529–30.

436. Manchanayake J, Chitturi S, Nolan C, Farrell GC. Postprandial hyperinsulinemia is universal in non-diabetic patients with nonalcoholic fatty liver disease. J Gastroenterol Hepatol 2011;26(3):510–6.

437. Gandhi CR, Sproat LA, Subbotin VM. Increased hepatic endothelin-1 levels and endothelin receptor density in cirrhotic rats. Life Sci 1995;58(1):55–62.

438. Keating SE, Hackett DA, George J, Johnson NA. Exercise and non-alcoholic fatty liver disease: A systematic review and meta-analysis. J Hepatol [Internet] 2012;57(1):157–66. Available from: http://www.sciencedirect.com/science/article/pii/S0168827812002103

439. Romero-Gomez M, Zelber-Sagi S, Trenell M. Treatment of NAFLD with diet, physical activity and exercise. J Hepatol 2017;67(4):829–46.

440. Younossi ZM, Blissett D, Blissett R, Henry L, Stepanova M, Younossi Y, Racila A, Hunt S, Beckerman R. The economic and clinical burden of nonalcoholic fatty liver disease in the United States and Europe. Hepatology. 2016 Nov;64(5):1577-86.

441. Mariotto AB, Yabroff KR, Shao Y, Feuer EJ, Brown ML. Projections of the cost of cancer care in the United States: 2010-2020. J Natl Cancer Inst [Internet] 2011;103(2):117–28. Available from: https://pubmed.ncbi.nlm.nih.gov/21228314

442. Miles A, Loughlin M, Polychronis A. Evidence-based healthcare, clinical knowledge and the rise of personalised medicine. Journal of Evaluation in Clinical Practice. 2008;

443. Burn J. Should we sequence everyone's genome? Yes. Bmj 2013;346.

444. Colditz GA. Overview of the epidemiology methods and applications: strengths and limitations of observational study designs. Crit Rev Food Sci Nutr 2010;50(S1):10–2.

445. Mariani AW, Pego-Fernandes PM. Observational studies: why are they so important? Sao Paulo Med J 2014;132(1):1–2.

446. Jackson SE, Chester JD. Personalised cancer medicine. Int J cancer 2015;137(2):262–6.

447. O'Gorman P, Naimimohasses S, Monaghan A, et al. Improvement in histological endpoints of MAFLD following a 12-week aerobic exercise intervention. Aliment Pharmacol Ther 2020;52(8):1387–98.

448. Åberg F, Färkkilä M. Drinking and obesity: Alcoholic liver disease/nonalcoholic fatty liver disease interactions. In: Seminars in liver disease. Thieme Medical Publishers; 2020. p. 154–62.

449. Vrabcová J, Svačinová K, Pechholdová M. Alcohol Consumption in Selected European Countries. In: Demography of Population Health, Aging and Health Expenditures. Springer; 2020. p. 187–200.

450. Pemmasani SK, Raman R, Mohapatra R, Vidyasagar M, Acharya A. A Review on the Challenges in Indian Genomics Research for Variant Identification and Interpretation [Internet]. Front. Genet. . 2020;11:753. Available from: https://www.frontiersin.org/article/10.3389/fgene.2020.00753

451. Jain A, Bhoyar RC, Pandhare K, et al. IndiGenomes: a comprehensive resource of genetic variants from over 1000 Indian genomes. Nucleic Acids Res [Internet] 2020;Available from: https://doi.org/10.1093/nar/gkaa923

452. Mahadevan L, Yesudas A, Sajesh PK, et al. Prevalence of genetic variants associated with cardiovascular disease risk and drug response in the Southern Indian population of Kerala. Indian J Hum Genet [Internet] 2014;20(2):175–84. Available from: https://pubmed.ncbi.nlm.nih.gov/25400347

453. Need AC, Goldstein DB. Next generation disparities in human genomics: concerns and remedies. Trends Genet [Internet] 2009;25(11):489–94. Available from: http://www.sciencedirect.com/science/article/pii/S0168952509001851